COMPOSITIONS OF ORTHOGONAL GLUTAMYL-tRNA AND AMINOACYL tRNA SYNTHETASE PAIRS AND USES THEREOF

5

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001]   This application claims priority to and benefit of Provisional Patent

Application USSN 60/485,451, filed July 7, 2003; and to Provisional Patent Application

10   USSN 60/513,109, filed October 20, 2003, the disclosures of which are incorporated herein

by reference in their entirety for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

15   [0002]   This invention was made with government support under Grant No. GM

62159 from the National Institutes of Health. The government may have certain rights to

this invention.

FIELD OF THE INVENTION

[0003]   The invention is in the field of translation biochemistry. The invention

20   relates to methods for producing and compositions of orthogonal glutamyl-tRNAs,

orthogonal glutamyl aminoacyl-tRNA synthetases and pairs thereof. The invention also

relates to methods of producing proteins in cells using such pairs and related compositions.

BACKGROUND OF THE INVENTION

[0004]   The genetic code of every known organism, from bacteria to humans,

25   encodes the same twenty common amino acids. Different combinations of the same twenty

natural amino acids form proteins that carry out virtually all the complex processes of life,

from photosynthesis to signal transduction and the immune response. In order to study and

modify protein structure and function, scientists have attempted to manipulate both the

genetic code and the amino acid sequence of proteins. However, it has been difficult to

30   remove the constraints imposed by the genetic code that limit proteins to twenty genetically

encoded standard building blocks (with the rare exception of selenocysteine (*see, e.g.,* A.

Bock et al., (1991), Molecular Microbiology 5:515-20) and pyrrolysine (*see, e.g.,* G. Srinivasan, et al., (2002), Science 296:1459-62).

[0005]   Some progress has been made to remove these constraints, although this progress has been limited, and the ability to rationally control protein structure and function

5   is still in its infancy. For example, chemists have developed methods and strategies to synthesize and manipulate the structures of small molecules (*see, e.g.,* E. J. Corey, & X.-M. Cheng, The Logic of Chemical Synthesis (Wiley-Interscience, New York, 1995)). Total synthesis (*see, e.g.,* B. Merrifield, (1986), Science 232:341-7 (1986)), and semi-synthetic methodologies (*see, e.g.,* D. Y. Jackson et al., (1994) Science 266:243-7; and, P. E.

10   Dawson, & S. B. Kent, (2000), Annual Review of Biochemistry 69:923-60), have made it possible to synthesize peptides and small proteins, but these methodologies have limited utility with proteins over 10 kilo Daltons (kDa). Mutagenesis methods, though powerful, are restricted to a limited number of structural changes. In a number of cases, it has been possible to competitively incorporate close structural analogues of common amino acids

15   throughout proteins. *See, e.g.,* R. Furter, (1998), Protein Science 7:419-26; K. Kirshenbaum, et al., (2002), ChemBioChem 3:235-7; and, V. Doring et al., (2001), Science 292:501-4.

[0006]   Early work demonstrated that the translational machinery of *E. coli* would accommodate amino acids similar in structure to the common twenty. *See* Hortin, G., and

20   Boime, I. (1983) Methods Enzymol. 96:777-784. This work was further extended by relaxing the specificity of endogenous *E. coli* synthetases so that they activate unnatural amino acids as well as their cognate natural amino acid. Moreover, it was shown that mutations in editing domains could also be used to extend the substrate scope of the endogenous synthetase. *See* Doring, V., et al., (2001) Science 292:501-504. However,

25   these strategies are limited to *recoding* the genetic code rather than *expanding* the genetic code and lead to varying degrees of substitution of one of the common twenty amino acids with an unnatural amino acid.

[0007]   Later it was shown that unnatural amino acids could be site-specifically incorporated into proteins *in vitro* by the addition of chemically aminoacylated orthogonal

30   amber suppressor tRNAs to an *in vitro* transcription/translation reaction. *See, e.g.,* Noren, C. J., et al. (1989) Science 244:182-188; Bain, J. D., et al., (1989) J. Am. Chem. Soc. 111:8013-8014; Dougherty, D. A. (2000) Curr. Opin. Chem. Biol. 4, 645-652; Cornish, V.

W., et al. (1995) Angew. Chem., Int. Ed. 34:621-633; J. A. Ellman, et al., (1992), Science
255:197-200; and, D. Mendel, et al., (1995), Annual Review of Biophysics and
Biomolecular Structure 24:435-462. These studies show that the ribosome and translation
factors are compatible with a large number of unnatural amino acids, even those with

5   unusual structures. Unfortunately, the chemical aminoacylation of tRNAs is difficult, and
the stoichiometric nature of this process severely limited the amount of protein that could be
generated.

[0008]   Unnatural amino acid tRNA complexes have been microinjected into cells.
For example, unnatural amino acids were introduced into the nicotinic acetylcholine

10   receptor in Xenopus oocytes (e.g., M.W. Nowak, et al. (1998), *In vivo incorporation of
unnatural amino acids into ion channels in Xenopus oocyte expression system*, Method
Enzymol. 293:504-529) by microinjection of a chemically misacylated Tetrahymena
thermophila tRNA (e.g., M.E. Saks, et al. (1996), *An engineered Tetrahymena tRNAGln for
in vivo incorporation of unnatural amino acids into proteins by nonsense suppression*, J.

15   Biol. Chem. 271:23169-23175), and the relevant mRNA. *See also* D.A. Dougherty (2000),
Unnatural amino acids as probes of protein structure and function, Curr. Opin. Chem. Biol.
4:645-652. Unfortunately, this methodology is limited to proteins in cells that can be
microinjected, and because the relevant tRNA is chemically acylated in vitro, and cannot be
re-acylated, the yields of protein are very low.

20   [0009]   To overcome these limitations, new components, e.g., orthogonal tRNAs,
orthogonal aminoacyl-tRNA synthetases and pairs thereof, were added to the protein
biosynthetic machinery of the prokaryote *Escherichia coli* (*E. coli*) (*see, e.g.,* L. Wang, et
al., (2001), Science 292:498-500), which allowed genetic encoding of unnatural amino acids
in vivo. A number of new amino acids with novel chemical, physical or biological

25   properties, including photoaffinity labels and photoisomerizable amino acids,
photocrosslinking amino acids (*see, e.g.,* Chin, J. W., et al. (2002) Proc. Natl. Acad. Sci. U.
S. A. 99:11020-11024; and, Chin, J. W., et al., (2002) J. Am. Chem. Soc. 124:9026-9027),
keto amino acids (*see, e.g.,* Wang, L., et al., (2003) Proc. Natl. Acad. Sci. U. S. A. 100:56-
61), heavy atom containing amino acids, and glycosylated amino acids have been

30   incorporated efficiently and with high fidelity into proteins in *E. coli* in response to, e.g., the
amber codon (TAG), using this methodology.

[0010]   Several other orthogonal pairs have been reported. Glutaminyl (*see, e.g.,* Liu, D. R., and Schultz, P. G. (1999) <u>Proc. Natl. Acad. Sci. U. S. A.</u> 96:4780-4785), aspartyl (*see, e.g.,* Pastrnak, M., et al., (2000) <u>Helv. Chim. Acta</u> 83:2277-2286), and tyrosyl (*see, e.g.,* Ohno, S., et al., (1998) <u>J. Biochem. (Tokyo, Jpn.)</u> 124:1065-1068; and, Kowal, A. K., et al., (2001) <u>Proc. Natl. Acad. Sci. U. S. A.</u> 98:2268-2273) systems derived from *S. cerevisiae* tRNAs and synthetases have been described for the potential incorporation of unnatural amino acids in *E. coli*. In addition, a leucine system, which includes an archaeal derived tRNA and a synthetase derived from, e.g., *Methanobacterium thermoautotrophicum*, has also been described. *See* Anderson and Schultz, (2003), <u>Biochemistry</u>, 42(32):9598-608. Systems derived from the *E. coli* glutaminyl (*see, e.g.,* Kowal, A. K., et al., (2001) <u>Proc. Natl. Acad. Sci. U. S. A.</u> 98:2268-2273) and tyrosyl (*see, e.g.,* Edwards, H., and Schimmel, P. (1990) <u>Mol. Cell. Biol.</u> 10:1633-1641) synthetase have been described for use in *S. cerevisiae*. The *E. coli* tyrosyl system has been used for the incorporation of 3-iodo-L-tyrosine *in vivo*, in mammalian cells. *See* Sakamoto, K., et al., (2002) <u>Nucleic Acids Res.</u> 30:4692-4699.

[0011]   Unnatural amino acids that have been incorporated into proteins, e.g., using variants of the tyrosyl-tRNA synthetase, are aryl derivatives, including *p*-azido- (*see, e.g.,* Chin et al., (2002) *Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli*, <u>J. Am. Chem. Soc.</u>, 124:9026-9027), *p*-benzoyl- (*see, e.g.,* Chin et al., (2002) *Addition of a photocrosslinking amino acid to the genetic code of Escherchia coli*, <u>PNAS, USA</u>, 99:11020-11024), *p*-amino- (*see, e.g.,* Santoro et al., (2002), *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity*, <u>Nat. Biotechnology</u>, 20:1044-1048), *p*-isopropyl- (*see, e.g.,* Santoro et al., (2002), *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity*, <u>Nat. Biotechnology</u>, 20:1044-1048), *m*-acetyl-(*see, e.g.,* Wang et al., (2003), *Addition of the keto functional group to the genetic code of Escherichia coli*, <u>PNAS, USA</u>, 100:56-61), and *p*-acetyl-phenylalanine (Wang et al., (2003), *Addition of the keto functional group to the genetic code of Escherichia coli*, <u>PNAS, USA</u>, 100:56-61); *O*-methyl- (*see, e.g.,* Wang et al., (2001), *Expanding the genetic cod of Escherichia coli*, <u>Science</u> 292:498-500) and *O*-allyl-tyrosine (*see, e.g.,* Santoro et al., (2002), *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity*, <u>Nat. Biotechnology</u>, 20:1044-1048; and Zhang et al., (2002) *The selective incorporation of alkenes into proteins in Escherichia coli*, <u>Angew Chem. Int. Ed. Engl.</u>, 41:2840-2842); 3-(2-

naphthyl)alanine (*see, e.g.,* Wang et al., (2002) *Adding L-3-(2-Naphthyl)alanine to the genetic code of E. coli,* J. Am. Chem. Soc., 124:1836-1837); and a *p*-propargyloxy phenylalanine (*see, e.g.,* Deiters et al., (2003) *Adding Amino Acids with Novel Reactivity to the Genetic Code of Saccharomyces Cerevisiae,* in press). To further expand the genetic

5   code and increase the diversity of unnatural amino acid structures that can be incorporated into proteins in a cell, there is a need to develop improved and/or additional components of the biosynthetic machinery, e.g., orthogonal tRNAs, orthogonal aminoacyl-tRNA synthetases and/or unique codons. This invention fulfills these and other needs, as will be apparent upon review of the following disclosure.

10   **SUMMARY OF THE INVENTION**
     [0012]   The invention provides compositions and methods of producing orthogonal components for incorporating a selected amino acid into a growing polypeptide chain in response to a selector codon, e.g., a stop codon, a nonsense codon, a four or more base codon, etc., e.g., in vivo. For example, the invention provides orthogonal glutamyl-tRNAs

15   (glutamyl O-tRNAs), orthogonal glutamyl aminoacyl-tRNA synthetases (glutamyl O-RSs) and pairs thereof. These pairs can be used to incorporate selected amino acids into growing polypeptide chains.

     [0013]   Typically, a glutamyl O-tRNA of the invention includes at least about, e.g., a 45%, a 50%, a 60%, a 75%, a 80%, or a 90% or more suppression efficiency in the presence

20   of a cognate synthetase in response to a selector codon, as compared to the glutamyl O-tRNA corresponding to a polynucleotide sequence as set forth in SEQ ID NO.: 67 (AE(GC) tRNA). A composition that includes a glutamyl O-tRNA can optionally further include an orthogonal glutamyl aminoacyl-tRNA synthetase (glutamyl O-RS), where the glutamyl O-RS preferentially aminoacylates the glutamyl O-tRNA with a selected amino acid. In one

25   embodiment, the suppression efficiency of the glutamyl O-RS and the glutamyl O-tRNA together is, e.g., 5 fold, 10 fold, 15 fold, 20 fold, 25 fold or more greater than the suppression efficiency of the glutamyl O-tRNA lacking the glutamyl O-RS.

     [0014]   A composition that includes a glutamyl O-tRNA can optionally include a cell (e.g., a non-eukaryotic cell, such as an *E. coli* cell and the like, or a eukaryotic cell), and/or

30   a translation system.

[0015]   A cell (e.g., a non-eukaryotic cell, or a eukaryotic cell) comprising a translation system is also provided by the invention, where the translation system includes an orthogonal glutamyl-tRNA (glutamyl O-tRNA); an orthogonal aminoacyl-glutamyl-tRNA synthetase (glutamyl O-RS); and, a first selected amino acid. Typically, the glutamyl O-tRNA comprises at least about, e.g., a 45%, a 50%, a 60%, a 75%, a 80%, or a 90% or more suppression efficiency in the presence of a cognate synthetase in response to a first selector codon as compared to a glutamyl O-tRNA comprising or encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 67 (AE(GC) tRNA). The glutamyl O-tRNA recognizes the first selector codon, and the glutamyl O-RS preferentially aminoacylates the glutamyl O-tRNA with the first selected amino acid in response to the selector codon. In one embodiment, the glutamyl O-tRNA comprises or is encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 67 (AE(GC)), or a complementary polynucleotide sequence thereof. In one aspect, the glutamyl O-tRNA comprises a G:C base pair at position 10:28. In one embodiment, the glutamyl O-RS comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 69 (Af), 73 (Mm), 75 (Mt), or 77 (Ph), or a conservative variation thereof.

[0016]   A cell of the invention can optionally further comprise an additional different O-tRNA/O-RS pair and a second selected amino acid, e.g., where this O-tRNA recognizes a second selector codon and this O-RS preferentially aminoacylates the O-tRNA with the second selected amino acid. Optionally, a cell of the invention includes a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, where the polynucleotide comprises a selector codon that is recognized by the glutamyl O-tRNA.

[0017]   In certain embodiments, a cell of the invention includes an *E. coli* cell that includes an orthogonal glutamyl tRNA (glutamyl O-tRNA), an orthogonal glutamyl aminoacyl-tRNA synthetase (glutamyl O-RS), a selected amino acid, and a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, where the polynucleotide comprises the selector codon that is recognized by the glutamyl O-tRNA. Typically, the glutamyl O-tRNA comprises at least about, e.g., a 45%, a 50%, a 60%, a 75%, a 80%, or a 90% or more suppression efficiency in the presence of a cognate synthetase in response to a selector codon as compared to the glutamyl O-tRNA comprising or encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 67 (AE(GC) tRNA), and the glutamyl

O-RS preferentially aminoacylates the glutamyl O-tRNA with a selected amino acid in response to the selector codon.

[0018]   In certain embodiments of the invention, a glutamyl O-tRNA of the invention comprises or is encoded by a polynucleotide sequence as set forth in SEQ ID
5   NO.: 67, or a complementary polynucleotide sequence thereof. In one embodiment, the glutamyl O-tRNA comprises a G:C base pair at position 10:28. In certain embodiments of the invention, a glutamyl O-RS comprises an amino acid sequence comprising any one of SEQ ID NO.: 69, 73, 75 and/or 77, or a conservative variation thereof. In one embodiment, the glutamyl O-RS or a portion thereof is encoded by a polynucleotide sequence comprising
10   any one of SEQ ID NO.: 68, 72, 74, or 76, or a complementary polynucleotide sequence thereof.

[0019]   The glutamyl O-tRNA and/or the glutamyl O-RS of the invention can be derived from any of a variety of organisms (e.g., eukaryotic and/or non-eukaryotic organisms). For example, the glutamyl O-tRNA can be derived from one or more archaeal
15   tRNAs and/or the glutamyl O-RS can be derived from, e.g., a non-eukaryotic organism ·
(e.g., an *Archaeoglobus fulgidus* (*Af*), a *Methanosarcina mazei* (*Mm*), a *Methanobacterium thermoautotrophicum* (*Mt*), a *Pyrococcus horikoshii* (*Ph*), etc.).

[0020]   Polynucleotides are also a feature of the invention. A polynucleotide of the invention includes an artificial (e.g., man-made, and not naturally occurring) polynucleotide
20   comprising a nucleotide sequence as set forth in SEQ ID NO.: 67, and/or is complementary to or that encodes a polynucleotide sequence of the above. A polynucleotide of the invention also includes a nucleic acid that hybridizes to a polynucleotide described above, under highly stringent conditions over substantially the entire length of the nucleic acid. A polynucleotide of the invention also includes a polynucleotide that is, e.g., at least 75%, at
25   least 80%, at least 90%, at least 95%, at least 98% or more identical to that of a naturally occurring glutamyl tRNA (but is other than a naturally occurring glutamyl tRNA) and comprises a G:C base pair at position 10:28. Artificial polynucleotides that are, e.g., at least 80%, at least 90%, at least 95%, at least 98% or more identical to any of the above and/or a polynucleotide comprising a conservative variation of any the above, are also included in
30   polynucleotides of the invention.

[0021]   Vectors comprising a polynucleotide of the invention are also a feature of the invention. For example, a vector of the invention can include a plasmid, a cosmid, a phage, a virus, an expression vector, and/or the like. A cell comprising a vector of the invention is also a feature of the invention.

5   [0022]   Methods of producing components of a glutamyl O-tRNA/O-RS pair are also features of the invention. Components produced by these methods are also a feature of the invention. For example, methods of producing at least one glutamyl tRNA that are orthogonal to a cell (glutamyl O-tRNA) include aligning a plurality of glutamyl tRNA sequences (e.g., derived from a species other than the host cell, derived from an archaeal

10   species, etc.); and, determining a consensus sequence (e.g., by using an alignment program, such as BLAST or pileup). A library of mutant glutamyl tRNAs is generated using the consensus sequence; and an anticodon loop of each member of the library of mutant glutamyl tRNAs is mutated to allow recognition of a selector codon, thereby providing a library of potential O-tRNAs. A first population of cells of a first species, where the cells

15   comprise a member of the library of potential O-tRNAs, is subjected to negative selection. The negative selection eliminates cells that comprise a member of the library of potential O-tRNAs that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is endogenous to the cell. This provides a pool of tRNAs that are orthogonal to the cell of the first species, thereby providing at least one glutamyl O-tRNA. An O-tRNA produced by the methods of

20   the invention is also provided.

[0023]   In certain embodiments, the methods further comprise mutating G10:U28 base pair to a G10:C28 base pair. In certain embodiments, the methods further comprise subjecting to positive selection a second population of cells of the first species, where the cells comprise a member of the pool of tRNAs that are orthogonal to the cell of the first

25   species, a cognate aminoacyl-tRNA synthetase, and a positive selection marker. Using the positive selection, cells are selected or screened for those cells that comprise a member of the pool of tRNAs that is aminoacylated by the cognate aminoacyl-tRNA synthetase and that shows a desired response in the presence of the positive selection marker, thereby providing an O-tRNA. In certain embodiments, the second population of cells comprises

30   cells that were not eliminated by the negative selection.

[0024]   Methods for identifying an orthogonal glutamyl-aminoacyl-tRNA synthetase for use with a glutamyl O-tRNA are also provided. For example, methods include

subjecting to selection a population of cells of a first species, where the cells each comprise:
1) a member of a plurality of aminoacyl-tRNA synthetases (RSs), (e.g., the plurality of RSs
can include mutant RSs, RSs derived from a species other than a first species or both mutant
RSs and RSs derived from a species other than a first species); 2) the orthogonal glutamyl
tRNA (O-tRNA) (e.g., from one or more species); and 3) a polynucleotide that encodes a
positive selection marker and comprises at least one selector codon. In certain
embodiments, the glutamyl O-tRNA displays at least about, e.g., a 45%, a 50%, a 60%, a
75%, a 80%, or a 90% or more suppression efficiency in the presence of a cognate
synthetase in response to a selector codon as compared to the glutamyl O-tRNA comprising
or encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 67 (AE(GC) tRNA).

[0025]   Cells (e.g., a host cell) are selected or screened for those that show an
enhancement in suppression efficiency compared to cells lacking or having a reduced
amount of the member of the plurality of RSs. These selected/screened cells comprise an
active RS that aminoacylates the O-tRNA. The level of aminoacylation (in vitro or in vivo)
by the active RS of a first set of tRNAs from the first species is compared to the level of
aminoacylation (in vitro or in vivo) by the active RS of a second set of tRNAs from a
second species (e.g., where the second species is other than the first species); where the
level of aminoacylation is determined by a detectable substance (e.g., a labeled amino acid).
The active RS that more efficiently aminoacylates the second set of tRNAs compared to the
first set of tRNAs is selected, thereby providing the glutamyl orthogonal aminoacyl-tRNA
synthetase, e.g., glutamyl O-RS, for use with the glutamyl O-tRNA. An orthogonal
aminoacyl-tRNA synthetase identified by the method is also a feature of the invention.

[0026]   Methods of producing a protein in a cell (e.g., a non-eukaryotic cell, such as
an E. coli cell or the like, or a eukaryotic cell) with a selected amino acid at a specified
position are also a feature of the invention. For example, a method includes growing, in an
appropriate medium, a cell, where the cell comprises a nucleic acid that comprises at least
one selector codon and encodes a protein, providing the selected amino acid, incorporating
the selected amino acid into the specified position in the protein during translation of the
nucleic acid with the at least one selector codon, thereby producing the protein. The cell
further comprises: a glutamyl-tRNA (glutamyl O-tRNA) that functions in the cell and
recognizes the selector codon; and, an orthogonal glutamyl aminoacyl-tRNA synthetase
(glutamyl O-RS) that preferentially aminoacylates the glutamyl-O-tRNA with the selected

amino acid. Typically, the glutamyl O-tRNA comprises at least about, e.g., a 45%, a 50%, a 60%, a 75%, a 80%, or a 90% or more suppression efficiency in the presence of a cognate synthetase in response to the selector codon as compared to the glutamyl O-tRNA comprising or encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 67

5   (AE(GC) tRNA). A protein produced by this method is also a feature of the invention.

DEFINITIONS

[0027]   Before describing the invention in detail, it is to be understood that this invention is not limited to particular biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing

10   particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a cell" includes a combination of two or more cells; reference to "bacteria" includes mixtures of bacteria, and the like.

15   [0028]   Unless defined herein and below in the reminder of the specification, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains.

[0029]   Orthogonal glutamyl-tRNA: As used herein, an orthogonal glutamyl-tRNA (glutamyl-O-tRNA) is a tRNA that is orthogonal to a translation system of interest, where

20   the tRNA is: (1) identical or substantially similar to a naturally occurring glutamyl tRNA, (2) derived from a naturally occurring glutamyl tRNA by natural or artificial mutagenesis, (3) derived by any process that takes a sequence of a wild-type or mutant glutamyl tRNA sequence of (1) or (2) into account, (4) homologous to a wild-type or mutant glutamyl tRNA; (5) homologous to any example tRNA that is designated as a substrate for a glutamyl

25   tRNA synthetase in TABLE 3, or (6) a conservative variant of any example tRNA that is designated as a substrate for a glutamyl tRNA synthetase in TABLE 3. The glutamyl tRNA can exist charged with a amino acid, or in an uncharged state. It is also to be understood that a "glutamyl-O-tRNA" optionally is charged (aminoacylated) by a cognate synthetase with an amino acid other than glutamic acid. Indeed, it will be appreciated that a glutamyl-

30   O-tRNA of the invention is advantageously used to insert essentially any amino acid,

whether natural or artificial, into a growing polypeptide, during translation, in response to a selector codon.

[0030]   Orthogonal glutamyl amino acid synthetase: As used herein, an orthogonal glutamyl amino acid synthetase (glutamyl-O-RS) is an enzyme that preferentially

5   aminoacylates the glutamyl-O-tRNA with an amino acid in a translation system of interest. The amino acid that the glutamyl-O-RS loads onto the glutamyl-O-tRNA can be any amino acid, whether natural or artificial, and is not limited herein. The synthetase is optionally the same as or homologous to a naturally occurring glutamyl amino acid synthetase, or the same as or homologous to a synthetase designated as a glutamyl-O-RS in TABLE 3. For

10   example, the glutamyl-O-RS can be a conservative variant of a glutamyl-O-RS of TABLE 3, and/or can be at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% or more identical in sequence to a glutamyl-O-RS of TABLE 3.

[0031]   Orthogonal: As used herein, the term "orthogonal" refers to a molecule (e.g., an orthogonal tRNA (O-tRNA) and/or an orthogonal aminoacyl tRNA synthetase (O-RS))

15   that functions with endogenous components of a cell with reduced efficiency as compared to a corresponding molecule that is endogenous to the cell or translation system, or that fails to function with endogenous components of the cell. In the context of tRNAs and aminoacyl-tRNA synthetases, orthogonal refers to an inability or reduced efficiency, e.g., less than 20 % efficiency, less than 10 % efficiency, less than 5 % efficiency, or less than

20   1% efficiency, of an orthogonal tRNA to function with an endogenous tRNA synthetase compared to an endogenous tRNA to function with the endogenous tRNA synthetase, or of an orthogonal aminoacyl-tRNA synthetase to function with an endogenous tRNA compared to an endogenous tRNA synthetase to function with the endogenous tRNA. The orthogonal molecule lacks a functionally normal endogenous complementary molecule in the cell. For

25   example, an orthogonal tRNA in a cell is aminoacylated by any endogenous RS of the cell with reduced or even zero efficiency, when compared to aminoacylation of an endogenous tRNA by the endogenous RS. In another example, an orthogonal RS aminoacylates any endogenous tRNA a cell of interest with reduced or even zero efficiency, as compared to aminoacylation of the endogenous tRNA by an endogenous RS. A second orthogonal

30   molecule can be introduced into the cell that function with the first orthogonal molecule. For example, an orthogonal tRNA/RS pair includes introduced complementary components that function together in the cell with an efficiency (e.g., 45 % efficiency, 50% efficiency,

60% efficiency, 70% efficiency, 75% efficiency, 80% efficiency, 90% efficiency, 95% efficiency, or 99% or more efficiency) as compared to that of a control, e.g., a corresponding tRNA/RS endogenous pair, or an active orthogonal pair (e.g., a tyrosyl orthogonal tRNA/RS pair).

5   [0032]   Cognate: The term "cognate" refers to components that function together, e.g., a glutamyl tRNA and a glutamyl aminoacyl-tRNA synthetase. The components can also be referred to as being complementary.

       [0033]   Preferentially aminoacylates: The term "preferentially aminoacylates" refers to an efficiency, e.g., 70 % efficiency, 75 % efficiency, 85% efficiency, 90% efficiency, 95
10   % efficiency, or 99% or more efficiency, at which an O-RS aminoacylates an O-tRNA with a selected amino acid, e.g., an unnatural amino acid, as compared to the O-RS aminoacylating a naturally occurring tRNA or a starting material used to generate the O-tRNA.

       [0034]   Selector codon: The term "selector codon" refers to codons recognized by
15   the O-tRNA in the translation process and typically not recognized by an endogenous tRNA. The O-tRNA anticodon loop recognizes the selector codon on the mRNA and incorporates its amino acid, e.g., a selected amino acid, such as an unnatural amino acid, at this site in the polypeptide. Selector codons can include, e.g., nonsense codons, such as, stop codons, e.g., amber, ochre, and opal codons; four or more base codons; rare codons;
20   codons derived from natural or unnatural base pairs and/or the like.

       [0035]   Suppressor tRNA: A suppressor tRNA is a tRNA that alters the reading of a messenger RNA (mRNA) in a given translation system, e.g., by providing a mechanism for incorporating an amino acid into a polypeptide chain in response to a selector codon. For example, a suppressor tRNA can read through, e.g., a stop codon, a four base codon, a rare
25   codon, etc.

       [0036]   Suppression activity: As used herein, the term "suppression activity" refers, in general, to the ability of a tRNA (e.g., a suppressor tRNA) to allow translational read-through of a codon (e.g. a selector codon that is an amber codon or a 4-or-more base codon) that would otherwise result in the termination of translation or mistranslation (e.g., frame-
30   shifting). Suppression activity of a suppressor tRNA can be expressed as a percentage of

translational read-through observed compared to a second suppressor tRNA, or as compared to a control system, e.g., a control system lacking an O-RS.

[0037]   The present invention provides various means by which suppression activity can be quantitated. Percent suppression of a particular O-tRNA and O-RS against a selector

5   codon (e.g., an amber codon) of interest refers to the percentage of activity of a given expressed test marker (e.g., LacZ), that includes a selector codon, in a nucleic acid encoding the expressed test marker, in a translation system of interest, where the translation system of interest includes an O-RS and an O-tRNA, as compared to a positive control construct, where the positive control lacks the O-tRNA, the O-RS and the selector codon. Thus, for

10   example, if an active positive control marker construct that lacks a selector codon has an observed activity of X in a given translation system, in units relevant to the marker assay at issue, then percent suppression of a test construct comprising the selector codon is the percentage of X that the test marker construct displays under essentially the same environmental conditions as the positive control marker was expressed under, except that

15   the test marker construct is expressed in a translation system that also includes the O-tRNA and the O-RS. Typically, the translation system expressing the test marker also includes an amino acid that is recognized by the O-RS and O-tRNA. Optionally, the percent suppression measurement can be refined by comparison of the test marker to a "background" or "negative" control marker construct, which includes the same selector

20   codon as the test marker, but in a system that does not include the O-tRNA, O-RS and/or relevant amino acid recognized by the O-tRNA and/or O-RS. This negative control is useful in normalizing percent suppression measurements to account for background signal effects from the marker in the translation system of interest.

[0038]   Suppression efficiency can be determined by any of a number of assays

25   known in the art. For example, a β-galactosidase reporter assay can be used, e.g., a derivatized lacZ plasmid (where the construct has a selector codon in the lacZ nucleic acid sequence) is introduced into cells from an appropriate organism (e.g., an organism where the orthogonal components can be used) along with plasmid comprising an O-tRNA of the invention. A cognate synthetase can also be introduced (either as a polypeptide or a

30   polynucleotide that encodes the cognate synthetase when expressed). The cells are grown in media to a desired density, e.g., to an OD_{600} of about 0.5, and β-galactosidase assays are performed, e.g., using the BetaFluor™ β-Galactosidase Assay Kit (Novagen). Percent

suppression can be calculated as the percentage of activity for a sample relative to a comparable control, e.g., the value observed from the derivatived lacZ construct, where the construct has a corresponding sense codon at desired position rather than a selector codon.

[0039]   Translation system: The term "translation system" refers to the components that incorporate an amino acid into a growing polypeptide chain (protein). Components of a translation system can include, e.g., ribosomes, tRNAs, synthetases, mRNA and the like. The O-tRNA and/or the O-RSs of the invention can be added to or be part of an *in vitro* or *in vivo* translation system, e.g., in a non-eukaryotic cell, e.g., a bacterium (such as *E. coli*), or in a eukaryotic cell, e.g., a yeast cell, a mammalian cell, a plant cell, an algae cell, a fungus cell, an insect cell, and/or the like.

[0040]   Selected amino acid: The term "selected amino acid" refers to any desired naturally occurring amino acid or unnatural amino acid. As used herein, the term "unnatural amino acid" refers to any amino acid, modified amino acid, and/or amino acid analogue that is not one of the 20 common naturally occurring amino acids or seleno cysteine or pyrrolysine.

[0041]   Derived from: As used herein, the term "derived from" refers to a component that is isolated from or made using a specified molecule or organism, or information from the specified molecule or organism.

[0042]   Positive selection or screening marker: As used herein, the term "positive selection or screening marker" refers to a marker that, when present, e.g., expressed, activated, or the like, results in identification of a cell, which comprise the trait, e.g., cells with the positive selection marker, from those without the trait.

[0043]   Negative selection or screening marker: As used herein, the term "negative selection or screening marker" refers to a marker that when present, e.g., expressed, activated, or the like, allows identification of a cell that does not comprise a specified property or trait (e.g., as compared to a cell that does possess the property or trait).

[0044]   Reporter: As used herein, the term "reporter" refers to a component that can be used to identify and/or select target components of a system of interest. For example, a reporter can include a protein, e.g., an enzyme, that confers antibiotic resistance or sensitivity (e.g., β-lactamase, chloramphenicol acetyltransferase (CAT), and the like), a fluorescent screening marker (e.g., green fluorescent protein (e.g., (GFP), YFP, EGFP, RFP,

etc.), a luminescent marker (e.g., a firefly luciferase protein), an affinity based screening marker, or positive or negative selectable marker genes such as lacZ, β-gal/lacZ (β-galactosidase), Adh (alcohol dehydrogenase), his3, ura3, leu2, lys2, or the like.

[0045]   Eukaryote: As used herein, the term "eukaryote" refers to organisms belonging to the phylogenetic domain Eucarya, such as animals (e.g., mammals, insects, reptiles, birds, etc.), ciliates, plants (e.g., monocots, dicots, algae, etc.), fungi, yeasts, flagellates, microsporidia, protists, etc.

[0046]   Non-eukaryote: As used herein, the term "non-eukaryote" refers to non-eukaryotic organisms. For example, a non-eukaryotic organism can belong to the Eubacteria (e.g., *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, etc.) phylogenetic domain, or the Archaea (e.g., *Methanococcus jannaschii (Mj)*, *Methanosarcina mazei (Mm)*, *Methanobacterium thermoautotrophicum (Mt)*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Halobacterium* such as *Haloferax volcanii* and *Halobacterium* species *NRC-1*, *Archaeoglobus fulgidus (Af)*, *Pyrococcus furiosus (Pf)*, *Pyrococcus horikoshii (Ph)*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus (Ss)*, *Sulfolobus tokodaii*, *Aeuropyrum pernix (Ap)*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, etc.) phylogenetic domains.

[0047]   Conservative variant: As used herein, the term "conservative variant" refers to a translation component, e.g., a conservative variant O-tRNA or a conservative variant O-RS, that functionally performs similar to a base component from which the conservative variant is based, e.g., an O-tRNA or O-RS, but having variations in the sequence. For example, an O-RS will aminoacylate a complementary O-tRNA or a conservative variant O-tRNA with a selected amino acid, e.g., an unnatural amino acid, although the O-tRNA and the conservative variant O-tRNA do not have the same sequence. The conservative variant can have, e.g., one variation, two variations, three variations, four variations, or five or more variations in sequence, as long as the conservative variant is complementary to the corresponding O-tRNA or O-RS.

[0048]   Selection or screening agent: As used herein, the term "selection or screening agent" refers to an agent that, when present, allows for a selection/screening of certain components from a population. For example, a selection or screening agent includes, but is not limited to, e.g., a nutrient, an antibiotic, a wavelength of light, an

antibody, an expressed polynucleotide, or the like. The selection agent can be varied, e.g., by concentration, intensity, etc.

[0049]   In response to: As used herein, in the context of translation with O-tRNA and O-RS components, the term "in response to" refers to the process in which a tRNA of the invention recognizes a selector codon and incorporates the selected amino acid, which is bound to tRNA, into the growing polypeptide chain.

[0050]   Exogenous or heterologous: As used herein, the terms "exogenous" or "heterologous" as applied to polynucleotides or polypeptides refers to molecules that have been artificially supplied to a biological system (e.g., a cell) and are not native to that particular biological system. The terms indicate that the relevant material originated from a source other than the naturally occurring source, or refers to molecules having a non-natural configuration, genetic location or arrangement of parts. The terms "exogenous" and "heterologous" are sometimes used interchangeably with "recombinant."

[0051]   In contrast, for example, a "native" or "endogenous" gene is a gene that does not contain nucleic acid elements encoded by sources other than the chromosome or other genetic element on which it is normally found in nature. An endogenous gene, transcript or polypeptide is encoded by its natural chromosomal locus, and not artificially supplied to the cell.

[0052]   Encode: As used herein, the term "encode" refers to any process whereby the information in a polymeric macromolecule or sequence string is used to direct the production of a second molecule or sequence string that is different from the first molecule or sequence string. As used herein, the term is used broadly, and can have a variety of applications. In one aspect, the term "encode" describes the process of semi-conservative DNA replication, where one strand of a double-stranded DNA molecule is used as a template to encode a newly synthesized complementary sister strand by a DNA-dependent DNA polymerase.

[0053]   In another aspect, the term "encode" refers to any process whereby the information in one molecule is used to direct the production of a second molecule that has a different chemical nature from the first molecule. For example, a DNA molecule can encode an RNA molecule (e.g., by the process of transcription incorporating a DNA-dependent RNA polymerase enzyme). Also, an RNA molecule can encode a polypeptide,

as in the process of translation. When used to describe the process of translation, the term

"encode" also extends to the triplet codon that encodes an amino acid. In some aspects, an

RNA molecule can encode a DNA molecule, e.g., by the process of reverse transcription

incorporating an RNA-dependent DNA polymerase. In another aspect, a DNA molecule

5   can encode a polypeptide, where it is understood that "encode" as used in that case

incorporates both the processes of transcription and translation.

BRIEF DESCRIPTION OF THE FIGURES

[0054]   Figure 1, Panels A and B schematically illustrate design of orthogonal

tRNAs. Panel A schematically illustrates consensus archaeal tRNA^{Glu} and tRNA^{Gln}

10   sequences. Non-canonical base pairs are circled; tRNA anticodons are boxed. Highlighted

bases correspond to positions at which two different bases predominate approximately

equally. K corresponds to bases G or U; Y corresponds to bases C or U; R corresponds to

bases A or G; S corresponds to bases C or G. Panel B schematically illustrates initial

candidate orthogonal tRNAs. Underlined bases correspond to positions that vary from the

15   consensus sequence.

[0055]   Figure 2 provides a histogram demonstrating the orthogonality of candidate

tRNAs in E. coli, e.g., using a β-lactamase amber suppression assay to measure the

ampicillin IC_{50} values for E. coli expressing the indicated tRNA.

[0056]   Figure 3 provides a histogram demonstrating aminoacylation of the AE(GU)

20   tRNA by orthogonal synthetase candidates in E. coli, e.g., using a β-lactamase amber

suppression assay to measure the ampicillin IC_{50}s of E. coli expressing the indicated

glutamyl-tRNA synthetase together with the AE(GU) tRNA. Ampicillin IC_{50} values reflect

amber suppression efficiency in E. coli.

[0057]   Figure 4, Panels A, B and C illustrate optimization of the AE(GU) tRNA by

25   selection. Panel A schematically illustrates sequence and secondary structure of the

AE(NN) tRNA library, which differs from the AE(GU) tRNA only at positions 10 and 28,

which are randomized. .Underlined bases correspond to positions that vary from the

consensus sequence; tRNA anticodons are boxed. Panel B schematically illustrates

sequence and secondary structure of the AE(GC) tRNA, which is identified as the most

30   active tRNA within the AE(NN) library. Panel C provides a histogram demonstrating

aminoacylation of the AE(GC) tRNA in E. coli, e.g., using a β-lactamase assay to measure

the ampicillin IC$_{50}$ values for *E. coli* expressing the indicated glutamyl-tRNA synthetases together with the AE(GC) tRNA or the AE(GC) tRNA with no synthetase.

[0058]   **Figure 5** provides a histogram demonstrating orthogonality of archaeal synthetases *in vitro*. For example, aminoacylation of whole tRNA from either *E. coli* (Ec) or
5   *Halobacterium sp. NRC-1* (halo) with [^{3}H]-glutamic acid is assayed *in vitro*.

DETAILED DESCRIPTION

[0059]   In order to add additional unnatural amino acids to the genetic code *in vivo*, "orthogonal pairs" of an aminoacyl-tRNA synthetase and a tRNA are needed that can function efficiently in the translational machinery. These pairs are orthogonal to a given
10   translation system's other elements, e.g., the endogenous RS and tRNA of the system. Desired characteristics of the orthogonal pairs include tRNA that decode or recognize only a specific new codon, e.g., a selector codon, that is not decoded by any endogenous tRNA, and aminoacyl-tRNA synthetases that preferentially aminoacylate (or charge) its cognate tRNA with only a specific selected amino acid, e.g., an unnatural amino acid. The O-tRNA
15   is also not typically aminoacylated by endogenous synthetases. For example, in *E. coli*, an orthogonal pair will include an aminoacyl-tRNA synthetase that does not significantly cross-react with any of the endogenous tRNA, which there are 40 in *E. coli*, and an orthogonal tRNA that is not significantly aminoacylated by any of the endogenous synthetases, e.g., of which there are 21 in *E. coli*.

20   [0060]   This invention provides compositions of and methods for identifying and producing additional orthogonal tRNA-aminoacyl-tRNA synthetase pairs, e.g., glutamyl O-tRNA/ glutamyl O-RS pairs. A glutamyl O-tRNA of the invention is capable of mediating incorporation of a selected amino acid into a protein that is encoded by a polynucleotide, which comprises a selector codon that is recognized by the glutamyl O-tRNA, e.g., in vivo.
25   The anticodon loop of the glutamyl O-tRNA recognizes the selector codon on an mRNA and incorporates its amino acid, e.g., a selected amino acid, such as an unnatural amino acid, at this site in the polypeptide. An orthogonal glutamyl aminoacyl-tRNA synthetase of the invention preferentially aminoacylates (or "charges") its glutamyl O-tRNA with only a specific selected amino acid.

ORTHOGONAL GLUTAMYL tRNA/ ORTHOGONAL GLUTAMYL AMINOACYL-
tRNA SYNTHETASES AND PAIRS THEREOF

[0061]   Other features of translation systems that are suitable for making proteins
that include one or more selected amino acids, e.g., an unnatural amino acid, are described
in Published International Applications WO 2002/086075, entitled "METHODS AND
COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-
tRNA SYNTHETASE PAIRS" and WO 2002/085923, entitled "IN VIVO
INCORPORATION OF UNNATURAL AMINO ACIDS." In addition, see International
Application Number PCT/US2004/011786, filed April 16, 2004. Each of these applications
is incorporated herein by reference in its entirety. These translation systems can be adopted
to the present invention by the inclusion of an O-RS/O-tRNA of the invention.

[0062]   Translation systems on the invention generally comprise cells (e.g., non-
eukaryotic cells, or eukaryotic cells) that include an orthogonal glutamyl tRNA (glutamyl
O-tRNA), an orthogonal glutamyl aminoacyl tRNA synthetase (glutamyl O-RS), and a
selected amino acid, where the glutamyl O-RS aminoacylates the glutamyl O-tRNA with
the selected amino acid. An orthogonal pair of the invention includes of a glutamyl O-
tRNA, e.g., a suppressor tRNA, a frameshift tRNA, or the like, and a glutamyl O-RS.
Individual components are also provided in the invention.

[0063]   The glutamyl O-tRNA herein recognizes a selector codon and includes at
least about, e.g., a 45%, a 50%, a 60%, a 75%, a 80%, or a 90% or more suppression
efficiency in the presence of a cognate synthetase in response to a selector codon as
compared to the glutamyl O-tRNA comprising or encoded by a polynucleotide sequence as
set forth in SEQ ID NO.: 67 (AE(GC) tRNA) in the presence of a cognate O-RS. The
cognate glutamyl O-RS aminoacylates the O-tRNA with a selected amino acid. The cell or
other translation system uses the components to incorporate the selected amino acid into a
growing polypeptide chain, e.g., via a nucleic acid that comprises a polynucleotide that
encodes a polypeptide of interest, where the polynucleotide comprises a selector codon that
is recognized by the glutamyl O-tRNA. In certain embodiments of the invention, a cell
includes an E. coli cell that includes an orthogonal glutamyl tRNA (glutamyl O-tRNA), an
orthogonal glutamyl aminoacyl-tRNA synthetase (glutamyl O-RS), a selected amino acid;
and, a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest,

where the polynucleotide comprises the selector codon that is recognized by the glutamyl O-tRNA. The translation system can also be an in vitro system.

[0064]   In one embodiment, the suppression efficiency of the glutamyl O-RS and the glutamyl O-tRNA together is about, e.g., 5 fold, 10 fold, 15 fold, 20 fold, or 25 fold or more

5   greater than the suppression efficiency of the glutamyl O-tRNA lacking the glutamyl O-RS. In one aspect, the suppression efficiency of the glutamyl O-RS and the glutamyl O-tRNA together is at least about, e.g., 35%, 40%, 45%, 50%, 60%, 75%, 80%, or 90% or more of the suppression efficiency of an orthogonal tyrosyl-tRNA synthetase pair derived from *Methanococcus jannaschii.*

10   [0065]   The invention also features multiple O-tRNA/O-RS pairs in a cell, which allows incorporation of more than one selected amino acid. For example, the cell can further include an additional different O-tRNA/O-RS pair and a second selected amino acid, where this additional O-tRNA recognizes a second selector codon and this additional O-RS preferentially aminoacylates the O-tRNA with the second selected amino acid. For

15   example, a cell, which includes a glutamyl O-tRNA/O-RS pair (where the glutamyl O-tRNA recognizes, e.g., an amber selector codon), can further comprise a second orthogonal pair, e.g., leucyl, lysyl, tyrosyl, etc., (where the second O-tRNA recognizes a different selector codon, e.g., an opal, four-base, or the like).

[0066]   The glutamyl O-tRNA and/or the glutamyl O-RS can be naturally occurring

20   or can be derived by mutation of a naturally occurring or consensus or hypothetical tRNA and/or RS, e.g., by generating libraries of tRNAs and/or libraries of RSs, from a variety of organisms. For example, one strategy of producing an orthogonal glutamyl tRNA/ glutamyl aminoacyl-tRNA synthetase pair involves importing a heterologous (to the host cell) tRNA/synthetase pair from, e.g., a source other than the host cell, or multiple sources, into

25   the host cell. The properties of the heterologous synthetase candidate include, e.g., that it does not significantly charge any host cell tRNA, and the properties of the heterologous tRNA candidate include, e.g., that it is not aminoacylated by any host cell synthetase. In addition, the heterologous tRNA is orthogonal to all host cell synthetases.

[0067]   A second strategy for generating an orthogonal pair involves generating

30   mutant libraries from which to screen and/or select a glutamyl O-tRNA or glutamyl O-RS. These strategies can also be combined.

[0068]   Archaea is a good source of orthogonal pairs for, e.g., *E. coli.* Previous studies revealed that several archaeal tRNAs are not recognized by *E. coli* aminoacyl-tRNA synthetases (*see, e.g.,* Kwok and Wong, (1980) *Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA*

5   *synthetases as phylogenetic probes,* Can. J. Biochem., 58:213-218). Two such synthetase-tRNA pairs derived from archaea, for tyrosine (*see, e.g.,* Wang et al., (2000) *A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins,* J. Am. Chem. Soc., 122:5010-5011) and leucine (*see, e.g.,* Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal*

10   *leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression,* Biochemistry, Aug 19; 42(32):9598-9608), were developed for use, e.g., in *E. coli.* Both exhibited high levels of orthogonality and amber suppression efficiency. Moreover, these archaeal synthetases and tRNAs are efficiently produced and processed in *E. coli.* Finally, the recent proliferation of genomic sequence data for archaeal species and the commercial availability

15   of their genomic DNA have facilitated the cloning of synthetases from these organisms, as well as the design of compatible tRNAs.

[0069]   In addition, archaeal glutamyl-tRNA synthetases and tRNAs are good sources for the production of additional orthogonal pairs for several reasons. Archaeal glutamyl-tRNA synthetases are expected to tolerate changes in the anticodon loop of their

20   substrate tRNAs since the enzyme naturally accommodates four different tRNA anticodons: UUC, CUC, UUG, and CUG. This is due to the fact that in archaeal species, glutamyl-tRNA synthetases must recognize and acylate both $tRNA^{Glu}$ and $tRNA^{Gln}$ (*see, e.g.,* Tumbula et al., (2000) *Domain-specific recruitment of amide amino acids for protein synthesis,* Nature 407:106-110). An archaeal glutamyl-tRNA synthetase can tolerate

25   variance within at least two positions of the anticodon of its tRNA substrate, which can permit efficient aminoacylation of a tRNA bearing, e.g., an amber-suppressing CUA anticodon. In addition, *in vitro* aminoacylation data (*see, e.g.,* Kwok and Wong (1980) *Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes,* Can. J. Biochem., 58:213-

30   218) predict that archaeal glutamic acid and glutamine tRNAs should be orthogonal in *E. coli.* Finally, the three dimensional structure of the glutamyl-tRNA synthetase from *Thermus thermophilus* is known (*see, e.g.,* Nureki et al., (1995) *Architectures of class-*

defining and specific domains of glutamyl-tRNA synthetase, Science 31: 1958-1965), which can facilitate the design of libraries of synthetase variants for use in directed evolution experiments. *See* the section entitled "Sources and Hosts," herein, for additional information.

5   Orthogonal tRNA (O-tRNA)

[0070]   An orthogonal glutamyl tRNA (glutamyl O-tRNA) mediates incorporation of a selected amino acid into a protein that is encoded by a polynucleotide that comprises a selector codon that is recognized by the glutamyl O-tRNA, e.g., in vivo or in vitro. Typically, a glutamyl O-tRNA of the invention includes at least about, e.g., a 45%, a 50%, a

10   60%, a 75%, a 80%, or a 90% or more suppression efficiency in the presence of a cognate synthetase in response to a selector codon as compared to the glutamyl O-tRNA comprising or encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 67 (AE(GC) tRNA) in the presence of a cognate RS.

[0071]   Suppression efficiency can be determined by any of a number of assays

15   known in the art. For example, a *β*-galactosidase reporter assay can be used, e.g., a derivatized lacZ plasmid (where the construct has a selector codon n the lacZ nucleic acid sequence) is introduced into cells from an appropriate organism (e.g., an organism where the orthogonal components can be used) along with plasmid comprising a O-tRNA of the invention. A cognate synthetase can also be introduced (either as a polypeptide or a

20   polynucleotide that encodes the cognate synthetase when expressed). The cells are grown in media to a desired density, e.g., to an OD_{600} of about 0.5, and β-galactosidase assays are performed, e.g., using the BetaFluor™ β-Galactosidase Assay Kit (Novagen). Percent suppression can be calculated as the percentage of activity for a sample relative to a comparable control, e.g., the value observed from the derivatived lacZ construct, where the

25   construct has a corresponding sense codon at desired position rather than a selector codon.

[0072]   An example of a glutamyl O-tRNA of the invention is a polynucleotide sequence comprising SEQ ID NO.: 67. *See* Table 3 and Example 2, herein, for sequences of exemplary O-tRNA and O-RS molecules. *See also,* the section entitled "Nucleic Acid and Polypeptide Sequence and Variants" herein. In the tRNA molecule, Thymine (T) is

30   replaced with Uracil (U). The glutamyl O-tRNA of the invention also optionally includes a G:C base pair at position 10:28. Additional modifications to the bases can also be present. The invention also includes conservative variations of glutamyl O-tRNA. For example,

conservative variations of glutamyl O-tRNA include those molecules that function like the glutamyl O-tRNA of SEQ ID NO.: 67 and maintain the tRNA L-shaped structure, but do not have the same sequence (and are other than wild type glutamyl tRNA molecules). *See also,* the section herein entitled "Nucleic acids and Polypeptides Sequence and Variants."

5   [0073]   The composition comprising a glutamyl O-tRNA can further include an orthogonal glutamyl aminoacyl-tRNA synthetase (glutamyl O-RS), where the glutamyl O-RS preferentially aminoacylates the glutamyl O-tRNA with a selected amino acid. In certain embodiments, a composition including a glutamyl O-tRNA can further include a translation system (e.g., in vitro or in vivo). A nucleic acid that comprises a polynucleotide

10   that encodes a polypeptide of interest, where the polynucleotide comprises a selector codon that is recognized by the glutamyl O-tRNA, or a combination of one or more of these can also be present in the cell. *See also* the section herein entitled "Orthogonal aminoacyl-tRNA synthetases."

[0074]   Methods of producing an orthogonal glutamyl tRNA (glutamyl O-tRNA) are

15   also a feature of the invention. A glutamyl O-tRNA produced by the method is also a feature of the invention. In certain embodiments of the invention, the glutamyl O-tRNAs can be produced by a consensus strategy. For example, potential orthogonal glutamyl tRNAs can be produced by aligning a plurality of glutamyl tRNA sequences (e.g., derived from a species other than a host cell's species, derived from an archaeal species, etc.);

20   determining a consensus sequence; and, generating a library of mutant glutamyl tRNAs using the consensus sequence.

[0075]   For example, a consensus sequence can be compiled with a computer program, e.g., the GCG program *pileup* or BLAST. Optionally, degenerate positions determined by the program are changed to the most frequent base at those positions. The

25   library of mutant tRNAs can be generated using various mutagenesis techniques known in the art. For example, the mutant tRNAs can be generated by site-specific mutations, random point mutations, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction or any combination thereof. In one embodiment of the invention, overlap extension of oligonucleotides in which each site of

30   the tRNA gene can be synthesized as a doped mixture of 90% the consensus sequence and 10% a mixture of the other 3 bases can be used to provide the library based on the consensus sequence. Other mixtures can also be used, e.g., 75% the consensus sequence

and 25% a mixture of the other 3 bases, 80% the consensus sequence and 20% a mixture of the other 3 bases, 95% the consensus sequence and 5% a mixture of the other 3 bases, etc.

[0076]   Additional mutations can be introduced at specific position(s), e.g., at nonconservative position(s), or at conservative position(s), at randomized position(s), or a combination of each in a desired loop or region of a tRNA, e.g., an anticodon loop, the acceptor stem, D arm or loop, variable loop, TψC arm or loop, other regions of the tRNA molecule, or a combination thereof. Typically, mutations in a glutamyl tRNA include mutating the anticodon loop of each member of the library of mutant glutamyl tRNAs to allow recognition of a selector codon. In one embodiment, mutations include a G10:C28 base pair. The method can further include adding an additional sequence (CCA) to a 3' terminus of the O-tRNA.

[0077]   Optionally, the secondary structure of a member of the library of potential glutamyl O-tRNAs can be analyzed to identify non-canonical base pairs in the secondary structure, and optionally methods of the invention include mutating the non-canonical base pairs (e.g., the non-canonical base pairs are mutated to canonical base pairs). Typically, a glutamyl O-tRNA possesses an improvement of orthogonality for a desired organism compared to the starting material, e.g., the plurality of tRNA sequences, while preserving its affinity towards a desired RS.

[0078]   The methods optionally include analyzing the homology of sequences of tRNAs and/or aminoacyl-tRNA synthetases to determine potential candidates for an O-tRNA, O-RS, and/or pairs thereof, that appear to be orthogonal for a specific organism. Computer programs known in the art and described herein can be used for the analysis. In one example, to choose potential orthogonal translational components for use in *E. coli*, a prokaryotic organism, a synthetase and/or a tRNA is chosen that does not display unusual homology to prokaryotic organisms.

[0079]   Typically, a glutamyl O-tRNA is obtained by subjecting to, e.g., negative selection, a population of cells of a first species, where the cells comprise a member of the plurality of potential glutamyl O-tRNAs. The negative selection eliminates cells that comprise a member of the library of potential glutamyl O-tRNAs that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is endogenous to the cell. This provides a pool of glutamyl tRNAs that are orthogonal to the cell of the first species.

[0080]   In certain embodiments, in the negative selection, a selector codon(s) is introduced into polynucleotide that encodes a negative selection marker, e.g., an enzyme that confers antibiotic resistance, e.g., β-lactamase, an enzyme that confers a detectable product, e.g., β-galactosidase, chloramphenicol acetyltransferase (CAT), e.g., a toxic product, such as barnase, at a nonessential position (e.g., still producing a functional barnase), etc. Screening/selection is optionally done by growing the population of cells in the presence of a selective agent (e.g., an antibiotic, such as ampicillin). In one embodiment, the concentration of the selection agent is varied.

[0081]   For example, to measure the activity of suppressor glutamyl tRNAs, a selection system is used that is based on the *in vivo* suppression of selector codon, e.g., nonsense or frameshift mutations introduced into a polynucleotide that encodes a negative selection marker, e.g., a gene for β-lactamase (*bla*). For example, polynucleotide variants, e.g., *bla* variants, with a selector codon at a certain position (e.g., A184), are constructed. Cells, e.g., bacteria, are transformed with these polynucleotides. In the case of an orthogonal glutamyl tRNA, which cannot be efficiently charged by endogenous *E. coli* synthetases, antibiotic resistance, e.g., ampicillin resistance, should be about or less than that for a bacteria transformed with no plasmid. If the glutamyl tRNA is not orthogonal, or if a heterologous synthetase capable of charging the tRNA is co-expressed in the system, a higher level of antibiotic, e.g., ampicillin, resistance is be observed. Cells, e.g., bacteria, are chosen that are unable to grow on LB agar plates with antibiotic concentrations about equal to cells transformed with no plasmids.

[0082]   In the case of a toxic product (e.g., ribonuclease or barnase), when a member of the plurality of potential glutamyl tRNAs is aminoacylated by endogenous host, e.g., *Escherichia coli* synthetases (i.e., it is not orthogonal to the host, e.g., *Escherichia coli* synthetases), the selector codon is suppressed and the toxic polynucleotide product produced leads to cell death. Cells harboring orthogonal glutamyl tRNAs or non-functional tRNAs survive.

[0083]   In one embodiment, the pool of glutamyl tRNAs that are orthogonal to a desired organism are then subjected to a positive selection in which a selector codon is placed in a positive selection marker, e.g., encoded by a drug resistance gene, such a β-lactamase gene. The positive selection is performed on a cell comprising a polynucleotide

encoding or comprising a member of the pool of tRNAs that are orthogonal to the cell, a polynucleotide encoding a positive selection marker, and a polynucleotide encoding a cognate RS. In certain embodiments, the second population of cells comprises cells that were not eliminated by the negative selection. The polynucleotides are expressed in the cell

5   and the cell is grown in the presence of a selection agent, e.g., ampicillin. Glutamyl tRNAs are then selected for their ability to be aminoacylated by the coexpressed cognate synthetase and to insert an amino acid in response to this selector codon. Typically, these cells show an enhancement in suppression efficiency compared to cells harboring non-functional tRNA(s), or tRNAs that cannot efficiently be recognized by the synthetase of interest. The

10   cell harboring the non-functional tRNA(s) or tRNAs that are not efficiently recognized by the synthetase of interest, are sensitive to the antibiotic. Therefore, glutamyl tRNAs that: (i) are not substrates for endogenous host, e.g., *Escherichia coli*, synthetases; (ii) can be aminoacylated by the synthetase of interest; and (iii) are functional in translation, survive both selections.

15   **[0084]**   The stringency of the selection, e.g., the positive selection, the negative selection or both the positive and negative selection, in the above described-methods, optionally includes varying the selection stringency. For example, because barnase is an extremely toxic protein, the stringency of the negative selection can be controlled by introducing different numbers of selector codons into the barnase gene and/or by using an

20   inducible promoter. In another example, the concentration of the selection or screening agent is varied (e.g., ampicillin concentration). In one aspect of the invention, the stringency is varied because the desired activity can be low during early rounds. Thus, less stringent selection criteria are applied in early rounds and more stringent criteria are applied in later rounds of selection. In certain embodiments, the negative selection, the positive

25   selection or both the negative and positive selection, can be repeated multiple times. Multiple different negative selection markers, positive selection markers or both negative and positive selection markers, can be used. In certain embodiments, the positive and negative selection marker can be the same.

[0085]   Other types of selections/screening can be used in the invention for

30   producing orthogonal translational components, e.g., a glutamyl O-tRNA, a glutamyl O-RS, and a glutamyl O-tRNA/O-RS pair. For example, the negative selection marker, the positive selection marker or both the positive and negative selection markers can include a

marker that fluoresces or catalyzes a luminescent reaction in the presence of a suitable

reactant. In another embodiment, a product of the marker is detected by fluorescence-

activated cell sorting (FACS) or by luminescence. Optionally, the marker includes an

affinity based screening marker. *See* Francisco, J. A., et al., (1993) *Production and*

5   *fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody*

fragment on the external surface. <u>Proc Natl Acad Sci U S A.</u> 90:10444-8.

[0086]   Additional general methods for producing a recombinant orthogonal tRNA

can be found, e.g., in International patent application WO 2002/086075, *supra*; and patent

applications USSN 60/496,548, USSN 60/479,931, and USSN 60/463,869, *supra. See also*

10   Forster et al., (2003) *Programming peptidomimetic synthetases by translating genetic codes*

designed de novo <u>PNAS</u> 100(11):6353-6357; and, Feng et al., (2003), *Expanding tRNA*

recognition of a tRNA synthetase by a single amino acid change, <u>PNAS</u> 100(10): 5676-

5681.

Orthogonal aminoacyl-tRNA synthetase (O-RS)

15   [0087]   A glutamyl O-RS of the invention preferentially aminoacylates a glutamyl

O-tRNA with a selected amino acid *in vitro* or *in vivo*. A glutamyl O-RS of the invention

can be provided to the translation system, e.g., a cell, by a polypeptide that includes a

glutamyl O-RS and/or by a polynucleotide that encodes a glutamyl O-RS or a portion

thereof. For example, a glutamyl O-RS, or a portion thereof, is encoded by a polynucleotide

20   sequence as set forth in any one of SEQ ID NO.: 68, 72, 74, or 76, or a complementary

polynucleotide sequence thereof. In another example, a glutamyl O-RS comprises an amino

acid sequence as set forth in any one of SEQ ID NO.: SEQ ID NO.: 69, 73, 75 and/or 77, or

a conservative variation thereof. *See, e.g.,* Table 3 and Example 2, herein, for sequences of

exemplary glutamyl O-RS molecules. *See also,* the section entitled "Nucleic Acid and

25   Polypeptide Sequence and Variants" herein.

[0088]   Methods for identifying an orthogonal glutamyl aminoacyl-tRNA synthetase

(O-RS), e.g., a glutamyl O-RS, for use with a glutamyl O-tRNA, are also a feature of the

invention. For example, a method includes subjecting to selection, e.g., positive selection, a

population of cells of a first species, where the cells individually comprise: 1) a member of

30   a plurality of aminoacyl-tRNA synthetases (RSs), (e.g., the plurality of RSs can include

mutant RSs, RSs derived from a species other than the first species or both mutant RSs and

RSs derived from a species other than the first species); 2) the orthogonal tRNA (O-tRNA)

(e.g., from one or more species); and 3) a polynucleotide that encodes a (e.g., positive) selection marker and comprises at least one selector codon. Cells are selected or screened for those that show an enhancement in suppression efficiency compared to cells lacking or with a reduced amount of the member of the plurality of glutamyl RSs. Suppression

5   efficiency can be measured by techniques known in the art and as described herein. Cells having an enhancement in suppression efficiency comprise an active glutamyl RS that aminoacylates the glutamyl O-tRNA. A level of aminoacylation (in vitro or in vivo) by the active RS of a first set of tRNAs from the first species is compared to the level of aminoacylation (in vitro or in vivo) by the active RS of a second set of tRNAs from the

10 second species. The level of aminoacylation can be determined by a detectable substance (e.g., a labeled amino acid or unnatural amino acid). The active RS that more efficiently aminoacylates the second set of tRNAs compared to the first set of tRNAs is typically selected, thereby providing an efficient (optimized) orthogonal glutamyl aminoacyl-tRNA synthetase for use with the glutamyl O-tRNA. A glutamyl O-RS, identified by the method is

15 also a feature of the invention.

[0089]   Any of a number of assays can be used to determine aminoacylation. These assays can be performed in vitro or in vivo. For example, in vitro aminoacylation assays are described in, e.g., Hoben and Soll (1985) Methods Enzymol. 113:55-59. Aminoacylation can also be determined by using a reporter along with orthogonal translation components

20 and detecting the reporter in a cell expressing a polynucleotide comprising at least one selector codon that encodes a protein. See also, International Publication No. WO 2002/085923, and, International Application Number PCT/US2004/011786.

[0090]   Identified glutamyl O-RS can be further manipulated to alter the substrate specificity of the synthetase, so that only a desired unnatural amino acid, but not any of the

25 common 20 amino acids are charged to the glutamyl O-tRNA. Methods to generate an orthogonal glutamyl aminoacyl tRNA synthetase with a substrate specificity for an unnatural amino acid include mutating the synthetase, e.g., at the active site in the synthetase, at the editing mechanism site in the synthetase, at different sites by combining different domains of synthetases, or the like, and applying a selection process. A strategy is

30 used, which is based on the combination of a positive selection followed by a negative selection. In the positive selection, suppression of the selector codon introduced at a nonessential position(s) of a positive marker allows cells to survive under positive selection

pressure. In the presence of both natural and unnatural amino acids, survivors thus encode active synthetases charging the orthogonal suppressor tRNA with either a natural or unnatural amino acid. In the negative selection, suppression of a selector codon introduced at a nonessential position(s) of a negative marker removes synthetases with natural amino

5   acid specificities. Survivors of the negative and positive selection encode synthetases that aminoacylate (charge) the orthogonal suppressor tRNA with unnatural amino acids only. These synthetases can then be subjected to further mutagenesis, e.g., DNA shuffling or other recursive mutagenesis methods.

[0091]   A library of mutant glutamyl O-RSs can be generated using various

10   mutagenesis techniques known in the art. For example, the mutant RSs can be generated by site-specific mutations, random point mutations, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction or any combination thereof. For example, a library of mutant glutamyl RSs can be produced from two or more other, e.g., smaller, less diverse "sub-libraries." Chimeric libraries of RSs are also included

15   in the invention. It should be noted that libraries of tRNA synthetases from various   ,
organism (e.g., microorganisms such as eubacteria or archaebacteria) such as libraries that comprise natural diversity (see, e.g., U.S. Patent No. 6,238,884 to Short et al; U.S. Patent No. 5,756,316 to Schallenberger et al; U.S. Patent No. 5,783,431 to Petersen et al; U.S. Patent No. 5,824,485 to Thompson et al; U.S. Patent No. 5,958,672 to Short et al), are

20   optionally constructed and screened for orthogonal pairs.

[0092]   Once the synthetases are subject to the positive and negative selection/screening strategy, these synthetases can then be subjected to further mutagenesis. For example, a nucleic acid that encodes the glutamyl O-RS can be isolated; a set of polynucleotides that encode mutated glutamyl O-RSs (e.g., by random mutagenesis, site-

25   specific mutagenesis, recombination or any combination thereof) can be generated from the nucleic acid; and, these individual steps or a combination of these steps can be repeated until a mutated glutamyl O-RS is obtained that preferentially aminoacylates the glutamyl O-tRNA with the unnatural amino acid. In one aspect of the invention, the steps are performed multiple times, e.g., at least two times.

30   [0093]   Additional levels of selection/screening stringency can also be used in the methods of the invention, for producing glutamyl O-tRNA, glutamyl O-RS, or pairs thereof. The selection or screening stringency can be varied on one or both steps of the method to

produce an O-RS. This could include, e.g., varying the amount of selection/screening agent that is used, etc. Additional rounds of positive and/or negative selections can also be performed. Selecting or screening can also comprise one or more of a change in amino acid permeability, a change in translation efficiency, a change in translational fidelity, etc.

5   Typically, the one or more change is based upon a mutation in one or more gene in an organism in which an orthogonal tRNA-tRNA synthetase pair is used to produce protein.

[0094]   Additional general information regarding generating O-RS, and altering the substrate specificity of the synthetase can be found in International Publication Number WO 2002/086075, and, International Application Number PCT/US2004/011786.

10   SOURCE AND HOST ORGANISMS

[0095]   The translational components of the invention can be derived from non-eukaryotic organisms. For example, the glutamyl O-tRNA can be derived from a non-eukaryotic organism (or a combination of organisms), e.g., an archaebacterium, such as *Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Halobacterium* such

15   as *Haloferax volcanii* and *Halobacterium* species *NRC-1, Archaeoglobus fulgidus, Pyrococcus furiosus, Pyrococcus horikoshii, Aeuropyrum pernix, Methanococcus maripaludis, Methanopyrus kandleri, Methanosarcina mazei (Mm), Pyrobaculum aerophilum, Pyrococcus abyssi, Sulfolobus solfataricus (Ss), Sulfolobus tokodaii, Thermoplasma acidophilum, Thermoplasma volcanium,* or the like, or a eubacterium, such

20   as *Escherichia coli, Thermus thermophilus, Bacillus stearothermophilus,* or the like, while the glutamyl O-RS can be derived from a non-eukaryotic organism (or a combination of organisms), e.g., an archaebacterium, such as *Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Halobacterium* such as *Haloferax volcanii* and *Halobacterium* species *NRC-1, Archaeoglobus fulgidus, Pyrococcus furiosus, Pyrococcus*

25   *horikoshii, Aeuropyrum pernix, Methanococcus maripaludis, Methanopyrus kandleri, Methanosarcina mazei, Pyrobaculum aerophilum, Pyrococcus abyssi, Sulfolobus solfataricus, Sulfolobus tokodaii, Thermoplasma acidophilum, Thermoplasma volcanium,* or the like, or a eubacterium, such as *Escherichia coli, Thermus thermophilus, Bacillus stearothermophilus,* or the like. In one embodiment, eukaryotic sources, e.g., plants, algae,

30   protists, fungi, yeasts, animals (e.g., mammals, insects, arthropods, etc.), or the like, can also be used as sources of glutamyl O-tRNAs and glutamyl O-RSs.

[0096]   The individual components of a glutamyl O-tRNA/O-RS pair can be derived from the same organism or different organisms. In one embodiment, the glutamyl O-tRNA/O-RS pair is from the same organism. Alternatively, the glutamyl O-tRNA and the glutamyl O-RS of the glutamyl O-tRNA/O-RS pair are from different organisms. For example, the glutamyl O-tRNA can be derived from, e.g., from a collection of archaeal organisms, and glutamyl O-RS can be derived from, e.g., an *Archaeoglobus fulgidus* (Af), a *Methanosarcina mazei* (Mm), a *Methanobacterium thermoautotrophicum* (Mt), a *Pyrococcus horikoshii* (Ph), etc.), a *Methanosarcina mazei* or the like.

[0097]   The glutamyl O-tRNA, glutamyl O-RS or glutamyl O-tRNA/O-RS pair can be selected or screened in vivo or in vitro and/or used in a cell, e.g., a non-eukaryotic cells, or eukaryotic cells, to produce a polypeptide with a selected amino acid. A non-eukaryotic cell can be from a variety of sources, e.g., a eubacterium, such as *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, or the like, or an archaebacterium, such as *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* such as *Haloferax volcanii* and *Halobacterium* species *NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei (Mm)*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus (Ss)*, *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, or the like. A eukaryotic cell can be from any of a variety of sources, e.g., a plant (e.g., complex plant such as monocots, or dicots), an algae, a protist, a fungus, a yeast (e.g., *Saccharomyces cerevisiae*), an animal (e.g., a mammal, an insect, an arthropod, etc.), or the like. Compositions of cells with translational components of the invention are also a feature of the invention.

[0098]   *See also,* International Application Number PCT/US2004/011786 for screening O-tRNA and/or O-RS in one species for use in another species.

SELECTOR CODONS

[0099]   Selector codons of the invention expand the genetic codon framework of protein biosynthetic machinery. For example, a selector codon includes, e.g., a unique three base codon, a nonsense codon, such as a stop codon, e.g., an amber codon (UAG), or an opal codon (UGA), an unnatural codon, at least a four base codon, a rare codon, or the like. A number of selector codons can be introduced into a desired gene, e.g., one or more, two

or more, more than three, etc. By using different selector codons, multiple orthogonal tRNA/synthetase pairs can be used that allow the simultaneous incorporation of multiple selected amino acids, e.g., unnatural amino acids, using these different selector codons.

[0100]   In one embodiment, the methods involve the use of a selector codon that is a stop codon for the incorporation of a selected amino acid in vivo in a cell. For example, a glutamyl O-tRNA is produced that recognizes the stop codon and is aminoacylated by a glutamyl O-RS with a selected amino acid. This glutamyl O-tRNA is not recognized by the naturally occurring host's aminoacyl-tRNA synthetases. Conventional site-directed mutagenesis can be used to introduce the stop codon at the site of interest in a polynucleotide encoding a polypeptide of interest. *See, e.g.,* Sayers, J.R., et al. (1988), *5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis.* Nucleic Acids Res, 791-802. When the glutamyl O-RS, glutamyl O-tRNA and the nucleic acid that encodes a polypeptide of interest are combined, e.g., in vivo, the selected amino acid is incorporated in response to the stop codon to give a polypeptide containing the selected amino acid at the specified position. In one embodiment of the invention, a stop codon used as a selector codon is an amber codon, UAG, and/or an opal codon, UGA. In one example, a genetic code in which UAG and UGA are both used as a selector codon can encode 22 amino acids while preserving the ochre nonsense codon, UAA, which is the most abundant termination signal.

[0101]   The incorporation of selected amino acids, *in vivo*, can be done without significant perturbation of the host cell. For example, in non-eukaryotic cells, such as *Escherichia coli*, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g., the amber suppressor tRNA, and the release factor 1 (RF1) (which binds to the UAG codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, e.g., either increasing the expression level of O-tRNA, e.g., the suppressor tRNA, or using an RF1 deficient strain. In eukaryotic cells, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g., the amber suppressor tRNA, and a eukaryotic release factor (e.g., eRF) (which binds to a stop codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, e.g., increasing the expression level of O-tRNA, e.g., the suppressor tRNA.

[0102]   Selected amino acids can also be encoded with rare codons. For example, when the arginine concentration in an in vitro protein synthesis reaction is reduced, the rare arginine codon, AGG, has proven to be efficient for insertion of Ala by a synthetic tRNA acylated with alanine. *See, e.g.,* Ma et al., Biochemistry, 32:7939 (1993). In this case, the synthetic tRNA competes with the naturally occurring tRNAArg, which exists as a minor species in *Escherichia coli*. In addition, some organisms do not use all triplet codons. An unassigned codon AGA in *Micrococcus luteus* has been utilized for insertion of amino acids in an in vitro transcription/translation extract. *See, e.g.,* Kowal and Oliver, Nucl. Acid. Res., 25:4685 (1997). Components of the invention can be generated to use these rare codons in vivo.

[0103]   Selector codons can also comprise extended codons, e.g., four or more base codons, such as, four, five, six or more base codons. Examples of four base codons include, e.g., AGGA, CUAG, UAGA, CCCU, and the like. Examples of five base codons include, e.g., AGGAC, CCCCU, CCCUC, CUAGA, CUACU, UAGGC, and the like. Methods of the invention include using extended codons based on frameshift suppression. Four or more base codons can insert, e.g., one or multiple selected amino acids, e.g., unnatural amino acids, into the same protein. In other embodiments, the anticodon loops can decode, e.g., at least a four-base codon, at least a five-base codon, or at least a six-base codon or more. Since there are 256 possible four-base codons, multiple unnatural amino acids can be encoded in the same cell using a four or more base codon. *See also,* Anderson et al., (2002) *Exploring the Limits of Codon and Anticodon Size,* Chemistry and Biology, 9:237-244; and, Magliery, (2001) *Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli,* J. Mol. Biol. 307: 755-769.

[0104]   For example, four-base codons have been used to incorporate unnatural amino acids into proteins using in vitro biosynthetic methods. *See, e.g.,* Ma et al., (1993) Biochemistry, 32:7939; and Hohsaka et al., (1999) J. Am. Chem. Soc., 121:34. CGGG and AGGU were used to simultaneously incorporate 2-naphthylalanine and an NBD derivative of lysine into streptavidin in vitro with two chemically acylated frameshift suppressor tRNAs. *See, e.g.,* Hohsaka et al., (1999) J. Am. Chem. Soc., 121:12194. In an in vivo study, Moore et al. examined the ability of tRNA^{Leu} derivatives with NCUA anticodons to suppress UAGN codons (N can be U, A, G, or C), and found that the quadruplet UAGA can

-33-

be decoded by a tRNA^{Leu} with a UCUA anticodon with an efficiency of 13 to 26% with little decoding in the 0 or −1 frame. *See* Moore et al., (2000) J. Mol. Biol., 298:195. In one embodiment, extended codons based on rare codons or nonsense codons can be used in invention, which can reduce missense readthrough and frameshift suppression at other unwanted sites.

[0105]   For a given system, a selector codon can also include one of the natural three base codons, where the endogenous system does not use (or rarely uses) the natural base codon. For example, this includes a system that is lacking a tRNA that recognizes the natural three base codon, and/or a system where the three base codon is a rare codon.

[0106]   Selector codons optionally include unnatural base pairs. These unnatural base pairs further expand the existing genetic alphabet. One extra base pair increases the number of triplet codons from 64 to 125. Properties of third base pairs include stable and selective base pairing, efficient enzymatic incorporation into DNA with high fidelity by a polymerase, and the efficient continued primer extension after synthesis of the nascent unnatural base pair. Descriptions of unnatural base pairs which can be adapted for methods and compositions include, e.g., Hirao, et al., (2002) *An unnatural base pair for incorporating amino acid analogues into protein*, Nature Biotechnology, 20:177-182. *See also* Wu, Y., et al., (2002) J. Am. Chem. Soc. 124:14626-14630. Other relevant publications are listed below.

[0107]   For in vivo usage, the unnatural nucleoside is membrane permeable and is phosphorylated to form the corresponding triphosphate. In addition, the increased genetic information is stable and not destroyed by cellular enzymes. Previous efforts by Benner and others took advantage of hydrogen bonding patterns that are different from those in canonical Watson-Crick pairs, the most noteworthy example of which is the iso-C:iso-G pair. *See, e.g.,* Switzer et al., (1989) J. Am. Chem. Soc., 111:8322; and Piccirilli et al., (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. These bases in general mispair to some degree with natural bases and cannot be enzymatically replicated. Kool and co-workers demonstrated that hydrophobic packing interactions between bases can replace hydrogen bonding to drive the formation of base pair. *See* Kool, (2000) Curr. Opin. Chem. Biol., 4:602; and Guckian and Kool, (1998) Angew. Chem. Int. Ed. Engl., 36, 2825. In an effort to develop an unnatural base pair satisfying all the above requirements, Schultz, Romesberg and co-workers have systematically synthesized and studied a series of

unnatural hydrophobic bases. A PICS:PICS self-pair is found to be more stable than natural base pairs, and can be efficiently incorporated into DNA by Klenow fragment of Escherichia coli DNA polymerase I (KF). *See, e.g.,* McMinn et al., (1999) J. Am. Chem. Soc., 121:11586; and Ogawa et al., (2000) J. Am. Chem. Soc., 122:3274. A 3MN:3MN

5   self-pair can be synthesized by KF with efficiency and selectivity sufficient for biological function. *See, e.g.,* Ogawa et al., (2000) J. Am. Chem. Soc., 122:8803. However, both bases act as a chain terminator for further replication. A mutant DNA polymerase has been recently evolved that can be used to replicate the PICS self pair. In addition, a 7AI self pair can be replicated. *See, e.g.,* Tae et al., (2001) J. Am. Chem. Soc., 123:7439. A novel

10 metallobase pair, Dipic:Py, has also been developed, which forms a stable pair upon binding Cu(II). *See* Meggers et al., (2000) J. Am. Chem. Soc., 122:10714. Because extended codons and unnatural codons are intrinsically orthogonal to natural codons, the methods of the invention can take advantage of this property to generate orthogonal tRNAs for them.

[0108]   A translational bypassing system can also be used to incorporate a selected

15 amino acid in a desired polypeptide. In a translational bypassing system, a large sequence is inserted into a gene but is not translated into protein. The sequence contains a structure that serves as a cue to induce the ribosome to hop over the sequence and resume translation downstream of the insertion.

SELECTED AMINO ACIDS

20 **[0109]**   As used herein, a selected amino acid refers to any desired naturally occurring amino acid or unnatural amino acid. A naturally occurring amino acid includes any one of the twenty genetically encoded alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine.

25 In one embodiment, the selected amino acid is incorporated into a growing polypeptide chain with high fidelity, e.g., at greater than 75% efficiency for a given selector codon, at greater than about 80% efficiency for a given selector codon, at greater than about 90% efficiency for a given selector codon, at greater than about 95% efficiency for a given selector codon, or at greater than about 99% or more efficiency for a given selector codon.

30 **[0110]**   As used herein, an unnatural amino acid refers to any amino acid, modified amino acid, or amino acid analogue other than selenocysteine and/or pyrrolysine and the

following twenty genetically encoded alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine. The generic structure of an alpha-amino acid is illustrated by Formula I:

I

$$H_2N \diagup \diagup C O_2H$$

R

5
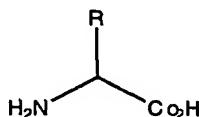
[0111]   An unnatural amino acid is typically any structure having Formula I wherein the R group is any substituent other than one used in the twenty natural amino acids. *See* e.g., <u>Biochemistry</u> by L. Stryer, 3rd ed. 1988, Freeman and Company, New York, for

10   structures of the twenty natural amino acids. Note that, the unnatural amino acids of the invention can be naturally occurring compounds other than the twenty alpha-amino acids above.

[0112]   Because the unnatural amino acids of the invention typically differ from the natural amino acids in side chain, the unnatural amino acids form amide bonds with other

15   amino acids, e.g., natural or unnatural, in the same manner in which they are formed in naturally occurring proteins. However, the unnatural amino acids have side chain groups that distinguish them from the natural amino acids. For example, R in Formula I optionally comprises an alkyl-, aryl-, acyl-, keto-, azido-, hydroxyl-, hydrazine, cyano-, halo-, hydrazide, alkenyl, alkynyl, ether, thiol, seleno-, sulfonyl-, borate, boronate, phospho,

20   phosphono, phosphine, heterocyclic, enone, imine, aldehyde, ester, thioacid, hydroxylamine, amine, and the like, or any combination thereof. Other unnatural amino acids of interest include, but are not limited to, amino acids comprising a photoactivatable cross-linker, spin-labeled amino acids, fluorescent amino acids, metal binding amino acids, metal-containing amino acids, radioactive amino acids, amino acids with novel functional

25   groups, amino acids that covalently or noncovalently interact with other molecules, photocaged and/or photoisomerizable amino acids, biotin or biotin-analogue containing amino acids, keto containing amino acids, amino acids comprising polyethylene glycol or polyether, heavy atom substituted amino acids, chemically cleavable or photocleavable amino acids, amino acids with an elongated side chain as compared to natural amino acids

(e.g., polyethers or long chain hydrocarbons, e.g., greater than about 5, greater than about
10 carbons, etc.), carbon-linked sugar-containing amino acids, amino thioacid containing
amino acids, and amino acids containing one or more toxic moiety. In some embodiments,
the unnatural amino acids have a photoactivatable cross-linker that is used, e.g., to link a
5   protein to a solid support. In one embodiment, the unnatural amino acids have a saccharide
moiety attached to the amino acid side chain (e.g., glycosylated amino acids) and/or other
carbohydrate modification.

[0113]   In addition to unnatural amino acids that contain novel side chains, unnatural
amino acids also optionally comprise modified backbone structures, e.g., as illustrated by
10   the structures of Formula II and III:

II
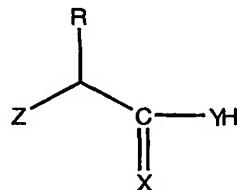
III
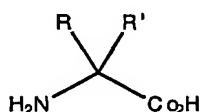
15   wherein Z typically comprises OH, NH_2, SH, NH-R', or S-R'; X and Y, which can be the
same or different, typically comprise S or O, and R and R', which are optionally the same or
different, are typically selected from the same list of constituents for the R group described
above for the unnatural amino acids having Formula I as well as hydrogen. For example,
unnatural amino acids of the invention optionally comprise substitutions in the amino or
20   carboxyl group as illustrated by Formulas II and III. Unnatural amino acids of this type
include, but are not limited to, α-hydroxy acids, α-thioacids α-aminothiocarboxylates, e.g.,
with side chains corresponding to the common twenty natural amino acids or unnatural side
chains. In addition, substitutions at the α-carbon optionally include L, D, or α-α-

disubstituted amino acids such as D-glutamate, D-alanine, D-methyl-O-tyrosine, aminobutyric acid, and the like. Other structural alternatives include cyclic amino acids, such as proline analogues as well as 3,4,6,7,8, and 9 membered ring proline analogues, β and γ amino acids such as substituted β-alanine and γ-amino butyric acid. Additional

5   unnatural amino acid structures of the invention include homo-beta-type structures, e.g., where there is, e.g., a methylene or amino group sandwiched adjacent to the alpha carbon, e.g., isomers of homo-beta-tyrosine, alpha-hydrazino-tyrosine. *See, e.g.,*
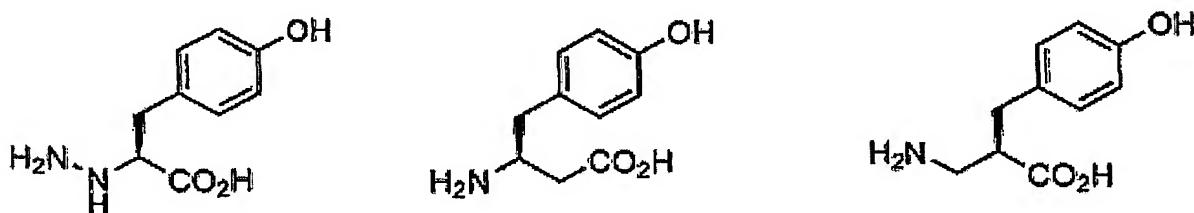
10   .

[0114]   Many unnatural amino acids are based on natural amino acids, such as glutamine, tyrosine, phenylalanine, and the like. For example, glutamine analogs of the invention include, but are not limited to, α-hydroxy derivatives, γ-substituted derivatives, cyclic derivatives, and amide substituted glutamine derivatives. Tyrosine analogs include

15   para-substituted tyrosines, ortho-substituted tyrosines, and meta substituted tyrosines, wherein the substituted tyrosine comprises an acetyl group, a benzoyl group, an amino group, a hydrazine, an hydroxyamine, a thiol group, a carboxy group, an isopropyl group, a methyl group, a C_6 - C_{20} straight chain or branched hydrocarbon, a saturated or unsaturated hydrocarbon, an O-methyl group, a polyether group, a nitro group, or the like. In addition,

20   multiply substituted aryl rings are also contemplated. Example phenylalanine analogs include, but are not limited to, para-substituted phenylalanines, ortho-substituted phenylalanines, and meta-substituted phenylalanines, wherein the substituent comprises a hydroxy group, a methoxy group, a methyl group, an allyl group, an aldehyde or keto group, or the like. Specific examples of unnatural amino acids include, but are not limited to, a *p*-

25   acetyl-L- phenylalanine, a *p*-propargyloxyphenylalanine, O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, a 3-methyl-phenylalanine, an O-4-allyl-L-tyrosine, a 4-propyl-L-tyrosine, a tri-O-acetyl-GlcNAcβ-serine, an L-Dopa, a fluorinated phenylalanine, an isopropyl-L-phenylalanine, a *p*-azido-L-phenylalanine, a *p*-acyl-L-phenylalanine, a *p*-benzoyl-L-

phenylalanine, an L-phosphoserine, a phosphonoserine, a phosphonotyrosine, a *p*-iodo-phenylalanine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, and an isopropyl-L-phenylalanine, and the like. The structures of a variety of unnatural amino acids are provided in, for example, Figures 16, 17, 18, 19, 26, and 29 of WO 2002/085923, *supra*.

                        Chemical Synthesis of Unnatural Amino Acids

[0115]   Many of the unnatural amino acids provided above are commercially available, e.g., from Sigma (USA) or Aldrich (Milwaukee, WI, USA). Those that are not commercially available are optionally synthesized as provided in various publications or using standard methods known to those of skill in the art. For organic synthesis techniques, *see, e.g.,* Organic Chemistry by Fessendon and Fessendon, (1982, Second Edition, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Third Edition, 1985, Wiley and Sons, New York); and Advanced Organic Chemistry by Carey and Sundberg (Third Edition, Parts A and B, 1990, Plenum Press, New York). Additional publications describing the synthesis of unnatural amino acids include, e.g., WO 2002/085923, *supra*; Matsoukas et al., (1995) J. Med. Chem., 38, 4660-4669; King, F.E. & Kidd, D.A.A. (1949) *A New Synthesis of Glutamine and of γ-Dipeptides of Glutamic Acid from Phthylated Intermediates*. J. Chem. Soc., 3315-3319; Friedman, O.M. & Chatterrji, R. (1959) *Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents.* J. Am. Chem. Soc. 81, 3750-3752; Craig, J.C. et al. (1988) *Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine).* J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmont, M. & Frappier, F. (1991) *Glutamine analogues as Potential Antimalarials,.* Eur. J. Med. Chem. 26, 201-5; Koskinen, A.M.P. & Rapoport, H. '(1989) *Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues.* J. Org. Chem. 54, 1859-1866; Christie, B.D. & Rapoport, H. (1985) *Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization.* J. Org. Chem. 1989:1859-1866; Barton et al., (1987) *Synthesis of Novel α-Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D-α-Amino-Adipic Acids, L-α-aminopimelic Acid and Appropriate Unsaturated Derivatives.* Tetrahedron Lett. 43:4297-4308; and, Subasinghe et al., (1992) *Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-*

sensitized site. J. Med. Chem. 35:4602-7. *See also,* International Application Number PCT/US03/41346, entitled "Protein Arrays," filed on December 22, 2003.

Cellular uptake of unnatural amino acids

[0116]   Unnatural amino acid uptake by a cell is one issue that is typically

5   considered when designing and selecting unnatural amino acids, e.g., for incorporation into a protein. For example, the high charge density of α-amino acids suggests that these compounds are unlikely to be cell permeable. Natural amino acids are taken up into the cell via a collection of protein-based transport systems often displaying varying degrees of amino acid specificity. A rapid screen can be done which assesses which unnatural amino

10   acids, if any, are taken up by cells. *See, e.g.,* International Application Number PCT/US03/41346, entitled "Protein Arrays," filed on December 22, 2003; and Liu, D.R. & Schultz, P.G. (1999) *Progress toward the evolution of an organism with an expanded genetic code.* PNAS United States 96:4780-4785. Although uptake is easily analyzed with various assays, an alternative to designing unnatural amino acids that are amenable to

15   cellular uptake pathways is to provide biosynthetic pathways to create amino acids *in vivo.*

Biosynthesis of Unnatural Amino Acids

[0117]   Many biosynthetic pathways already exist in cells for the production of amino acids and other compounds. While a biosynthetic method for a particular unnatural amino acid may not exist in nature, e.g., in a cell, the invention provides such methods. For

20   example, biosynthetic pathways for unnatural amino acids are optionally generated in a host cell by adding new enzymes or modifying existing host cell pathways. Additional new enzymes are optionally naturally occurring enzymes or artificially evolved enzymes. For example, the biosynthesis of *p*-aminophenylalanine (as presented in an example in WO 2002/085923, *supra*) relies on the addition of a combination of known enzymes from other

25   organisms. The genes for these enzymes can be introduced into a cell by transforming the cell with a plasmid comprising the genes. The genes, when expressed in the cell, provide an enzymatic pathway to synthesize the desired compound. Examples of the types of enzymes that are optionally added are provided in the examples below. Additional enzyme sequences are found, e.g., in Genbank. Artificially evolved enzymes are also optionally

30   added into a cell in the same manner. In this manner, the cellular machinery and resources of a cell are manipulated to produce unnatural amino acids.

[0118]   Indeed, any of a variety of methods can be used for producing novel enzymes for use in biosynthetic pathways, or for evolution of existing pathways, for the production of unnatural amino acids, in vitro or in vivo. Many available methods of evolving enzymes and other biosynthetic pathway components can be applied to the present invention to produce unnatural amino acids (or, indeed, to evolve synthetases to have new substrate specificities or other activities of interest). For example, DNA shuffling is optionally used to develop novel enzymes and/or pathways of such enzymes for the production of unnatural amino acids (or production of new synthetases), in vitro or in vivo. *See, e.g.,* Stemmer (1994), *Rapid evolution of a protein in vitro by DNA shuffling,* Nature 370(4):389-391; and, Stemmer, (1994), *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution,* Proc. Natl. Acad. Sci. USA., 91:10747-10751. A related approach shuffles families of related (e.g., homologous) genes to quickly evolve enzymes with desired characteristics. An example of such "family gene shuffling" methods is found in Crameri et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature, 391(6664): 288-291. New enzymes (whether biosynthetic pathway components or synthetases) can also be generated using a DNA recombination procedure known as "incremental truncation for the creation of hybrid enzymes" ("ITCHY"), e.g., as described in Ostermeier et al. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" Nature Biotech 17:1205. This approach can also be used to generate a library of enzyme or other pathway variants which can serve as substrates for one or more in vitro or in vivo recombination methods. See also, Ostermeier et al. (1999) "Combinatorial Protein Engineering by Incremental Truncation," Proc. Natl. Acad. Sci. USA, 96: 3562-67, and Ostermeier et al. (1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts," Biological and Medicinal Chemistry, 7: 2139-44. Another approach uses exponential ensemble mutagenesis to produce libraries of enzyme or other pathway variants that are, e.g., selected for an ability to catalyze a biosynthetic reaction relevant to producing an unnatural amino acid (or a new synthetase). In this approach, small groups of residues in a sequence of interest are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Examples of such procedures, which can be adapted to the present invention to produce new enzymes for the production of unnatural amino acids (or new synthetases) are found in Delegrave & Youvan (1993) Biotechnology Research 11:1548-1552. In yet another approach, random or semi-random mutagenesis

using doped or degenerate oligonucleotides for enzyme and/or pathway component engineering can be used, e.g., by using the general mutagenesis methods of e.g., Arkin and Youvan (1992) "Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis" Biotechnology 10:297-300; or Reidhaar-Olson et al. (1991)

5   "Random mutagenesis of protein sequences using oligonucleotide cassettes" Methods Enzymol. 208:564-86. Yet another approach, often termed a "non-stochastic" mutagenesis, which uses polynucleotide reassembly and site-saturation mutagenesis can be used to produce enzymes and/or pathway components, which can then be screened for an ability to perform one or more synthetase or biosynthetic pathway function (e.g., for the production of

10   unnatural amino acids in vivo). See, e.g., Short "Non-Stochastic Generation of Genetic Vaccines and Enzymes" WO 00/46344.

[0119]   An alternative to such mutational methods involves recombining entire genomes of organisms and selecting resulting progeny for particular pathway functions (often referred to as "whole genome shuffling"). This approach can be applied to the

15   present invention, e.g., by genomic recombination and selection of an organism (e.g., an *E. coli* or other cell) for an ability to produce an unnatural amino acid (or intermediate thereof). For example, methods taught in the following publications can be applied to pathway design for the evolution of existing and/or new pathways in cells to produce unnatural amino acids in vivo: Patnaik et al. (2002) "Genome shuffling of lactobacillus for

20   improved acid tolerance" Nature Biotechnology, 20(7): 707-712; and Zhang et al. (2002) "Genome shuffling leads to rapid phenotypic improvement in bacteria" Nature, February 7, 415(6872): 644-646.

[0120]   Other techniques for organism and metabolic pathway engineering, e.g., for the production of desired compounds are also available and can also be applied to the

25   production of unnatural amino acids. Examples of publications teaching useful pathway engineering approaches include: Nakamura and White (2003) "Metabolic engineering for the microbial production of 1,3 propanediol" Curr. Opin. Biotechnol. 14(5):454-9; Berry et al. (2002) "Application of Metabolic Engineering to improve both the production and use of Biotech Indigo" J. Industrial Microbiology and Biotechnology 28:127-133; Banta et al.

30   (2002) "Optimizing an artificial metabolic pathway: Engineering the cofactor specificity of Corynebacterium 2,5-diketo-D-gluconic acid reductase for use in vitamin C biosynthesis"

Biochemistry, 41(20), 6226-36; Selivonova et al. (2001) "Rapid Evolution of Novel Traits in Microorganisms" Applied and Environmental Microbiology, 67:3645, and many others.

[0121]   Regardless of the method used, typically, the unnatural amino acid produced with an engineered biosynthetic pathway of the invention is produced in a concentration sufficient for efficient protein biosynthesis, e.g., a natural cellular amount, but not to such a degree as to significantly affect the concentration of other cellular amino acids or to exhaust cellular resources. Typical concentrations produced *in vivo* in this manner are about 10 mM to about 0.05 mM. Once a cell is engineered to produce enzymes desired for a specific pathway and an unnatural amino acid is generated, *in vivo* selections are optionally used to further optimize the production of the unnatural amino acid for both ribosomal protein synthesis and cell growth.

NUCLEIC ACID AND POLYPEPTIDE SEQUENCE AND VARIANTS

[0122]   As described above and below, the invention provides for nucleic acid polynucleotide sequences, e.g., glutamyl O-tRNAs and glutamyl O-RSs, and polypeptide amino acid sequences, e.g., glutamyl O-RSs, and, e.g., compositions, systems and methods comprising said sequences. Examples of said sequences, e.g., glutamyl O-tRNAs and glutamyl O-RSs are disclosed herein (*see*, e.g., SEQ ID NO. 67, 68, 72, 74, 76, 69, 73, 75 and/or 77, along with others listed in Table 3 herein). However, one of skill in the art will appreciate that the invention is not limited to those sequences disclosed herein, e.g., as in the Examples. One of skill will appreciate that the invention also provides many related and even unrelated sequences with the functions described herein, e.g., encoding a glutamyl O-tRNA or a glutamyl O-RS.

[0123]   The invention provides polypeptides (glutamyl O-RSs) and polynucleotides, e.g., glutamyl O-tRNA, polynucleotides that encode glutamyl O-RSs or portions thereof, oligonucleotides used to isolate aminoacyl-tRNA synthetase clones, etc. Polynucleotides of the invention include those that encode proteins or polypeptides of interest of the invention with one or more selector codon. In addition, polynucleotides of the invention include, e.g., a polynucleotide comprising a nucleotide sequence comprising SEQ ID NO.: 67; a polynucleotide that is complementary to or that encodes a polynucleotide sequence thereof. A polynucleotide of the invention also includes a polynucleotide that encodes a polypeptide of the invention. Similarly, an artificial nucleic acid that hybridizes to a polynucleotide

indicated above under highly stringent conditions over substantially the entire length of the nucleic acid (and is other than a naturally polynucleotide) is a polynucleotide of the invention. In one embodiment, a composition includes a polypeptide of the invention and an excipient (e.g., buffer, water, pharmaceutically acceptable excipient, etc.).

[0124]   The invention also provides an antibody or antisera specifically immunoreactive with a polypeptide of the invention. A polynucleotide of the invention also includes an artificial polynucleotide that is, e.g., at least 75%, at least 80%, at least 90%, at least 95%, at least 98% or more identical to that of a naturally occurring glutamyl tRNA, (but is other than a naturally occurring glutamyl tRNA. An artificial polynucleotide or polypeptide is a polynucleotide or a polypeptide that is man made (e.g., via recombinant expression or chemical synthesis) and is not naturally occurring.

[0125]   In certain embodiments, a vector (e.g., a plasmid, a cosmid, a phage, a virus, etc.) comprises a polynucleotide of the invention. In one embodiment, the vector is an expression vector. In another embodiment, the expression vector includes a promoter operably linked to one or more of the polynucleotides of the invention. In another embodiment, a cell comprises a vector that includes a polynucleotide of the invention.

[0126]   One of skill will also appreciate that many variants of the disclosed sequences are included in the invention. For example, conservative variations of the disclosed sequences that yield a functionally identical sequence are included in the invention. Variants of the nucleic acid polynucleotide sequences, wherein the variants hybridize to at least one disclosed sequence, are considered to be included in the invention. Unique subsequences of the sequences disclosed herein, as determined by, e.g., standard sequence comparison techniques, are also included in the invention.

Conservative variations

[0127]   Owing to the degeneracy of the genetic code, "silent substitutions" (i.e., substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid. Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

[0128]   "Conservative variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or, where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 4%, 2% or 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid. Thus, "conservative variations" of a listed polypeptide sequence of the present invention include substitutions of a small percentage, typically less than 5%, more typically less than 2% or 1%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. Finally, the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional sequence, is a conservative variation of the basic nucleic acid.

[0129]   Conservative substitution tables providing functionally similar amino acids are well known in the art. The following sets forth example groups which contain natural amino acids that include "conservative substitutions" for one another.

Conservative Substitution Groups

Nonpolar and/or Aliphatic Side Chains	Polar, Uncharged Side Chains	Aromatic Side Chains	Positively Charged Side Chains	Negatively Charged Side Chains
Glycine Alanine Valine Leucine Isoleucine Proline	Serine Threonine Cysteine Methionine Asparagine Glutamine	Phenylalanine Tyrosine Tryptophan	Lysine Arginine Histidine	Aspartate Glutamate

Nucleic Acid Hybridization

[0130]   Comparative hybridization can be used to identify nucleic acids of the invention, such as SEQ ID NO.: 67, including conservative variations of nucleic acids of the invention, and this comparative hybridization method is a preferred method of distinguishing nucleic acids of the invention. In addition, target nucleic acids which hybridize to a nucleic acid represented by SEQ ID NO: 67 under high, ultra-high and ultra-

ultra high stringency conditions are a feature of the invention. Examples of such nucleic acids include those with one or a few silent or conservative nucleic acid substitutions as compared to a given nucleic acid sequence.

[0131]   A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least ½ as well to the probe as to the perfectly matched complementary target, i.e., with a signal to noise ratio at least ½ as high as hybridization of the probe to the target under conditions in which the perfectly matched probe binds to the perfectly matched complementary target with a signal to noise ratio that is at least about 5x-10x as high as that observed for hybridization to any of the unmatched target nucleic acids.

[0132]   Nucleic acids "hybridize" when they associate, typically in solution. Nucleic acids hybridize due to a variety of well characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part I chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), as well as in Ausubel, *supra*. Hames and Higgins (1995) Gene Probes 1 IRL Press at Oxford University Press, Oxford, England, (Hames and Higgins 1) and Hames and Higgins (1995) Gene Probes 2 IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides.

[0133]   An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see,* Sambrook, *supra* for a description of SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove background probe signal. An example low stringency wash is 2x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 5x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

[0134]   "Stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), *supra.* and in Hames and

5   Higgins, 1 and 2. Stringent hybridization and wash conditions can easily be determined empirically for any test nucleic acid. For example, in determining stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased (e.g., by increasing temperature, decreasing salt concentration, increasing detergent concentration and/or increasing the concentration of organic solvents such as formalin in the hybridization

10   or wash), until a selected set of criteria are met. For example, in highly stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased until a probe binds to a perfectly matched complementary target with a signal to noise ratio that is at least 5x as high as that observed for hybridization of the probe to an unmatched target.

15   [0135]   "Very stringent" conditions are selected to be equal to the thermal melting point (T_m) for a particular probe. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the test sequence hybridizes to a perfectly matched probe. For the purposes of the present invention, generally, "highly stringent" hybridization and wash conditions are selected to be about 5° C lower than the T_m for the specific sequence at a

20   defined ionic strength and pH.

[0136]   "Ultra high-stringency" hybridization and wash conditions are those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x as high as that observed for hybridization to any of the unmatched target nucleic

25   acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.

[0137]   Similarly, even higher levels of stringency can be determined by gradually increasing the hybridization and/or wash conditions of the relevant hybridization assay. For

30   example, those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x, 20X, 50X, 100X, or 500X or more as high

as that observed for hybridization to any of the unmatched target nucleic acids. A target
nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio
of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind
to the probe under ultra-ultra-high stringency conditions.

5   [0138]   Nucleic acids which do not hybridize to each other under stringent conditions
are still substantially identical if the polypeptides which they encode are substantially
identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum
codon degeneracy permitted by the genetic code.

Unique subsequences

10 [0139]   In one aspect, the invention provides a nucleic acid that comprises a unique
subsequence in a nucleic acid selected from the sequences of glutamyl O-tRNAs and
glutamyl O-RSs disclosed herein. The unique subsequence is unique as compared to a
nucleic acid corresponding to any known glutamyl O-tRNA or glutamyl O-RS nucleic acid
sequence. Alignment can be performed using, e.g., BLAST set to default parameters. Any
15 unique subsequence is useful, e.g., as a probe to identify the nucleic acids of the invention.

[0140]   Similarly, the invention includes a polypeptide which comprises a unique
subsequence in a polypeptide selected from the sequences of glutamyl O-RSs disclosed
herein. Here, the unique subsequence is unique as compared to a polypeptide corresponding
to any of known polypeptide sequence.

20 [0141]   The invention also provides for target nucleic acids which hybridizes under
stringent conditions to a unique coding oligonucleotide which encodes a unique
subsequence in a polypeptide selected from the sequences of glutamyl O-RSs wherein the
unique subsequence is unique as compared to a polypeptide corresponding to any of the
control polypeptides (e.g., parental sequences from which synthetases of the invention were
25 derived, e.g., by mutation). Unique sequences are determined as noted above.

Sequence comparison, identity, and homology

[0142]   The terms "identical" or percent "identity," in the context of two or more
nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that
are the same or have a specified percentage of amino acid residues or nucleotides that are
30 the same, when compared and aligned for maximum correspondence, as measured using

one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

[0143]   The phrase "substantially identical," in the context of two nucleic acids or polypeptides (e.g., DNAs encoding an glutamyl O-tRNA or glutamyl O-RS, or the amino acid sequence of an O-RS) refers to two or more sequences or subsequences that have at least about 60%, about 80%, about 90-95%, about 98%, about 99% or more nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry. Preferably, the "substantial identity" exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably, the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

[0144]   Proteins and/or protein sequences are "homologous" when they are derived, naturally or artificially, from a common ancestral protein or protein sequence. Similarly, nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. For example, any naturally occurring nucleic acid can be modified by any available mutagenesis method to include one or more selector codon. When expressed, this mutagenized nucleic acid encodes a polypeptide comprising one or more selected amino acid, e.g. unnatural amino acid. The mutation process can, of course, additionally alter one or more standard codon, thereby changing one or more standard amino acid in the resulting mutant protein as well. Homology is generally inferred from sequence similarity between two or more nucleic acids or proteins (or sequences thereof). The precise percentage of similarity between sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity is routinely used to establish homology. Higher levels of sequence similarity, e.g., 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% or more, can also be used to establish homology. Methods for determining sequence similarity percentages (e.g., BLASTP and BLASTN using default parameters) are described herein and are generally available.

[0145]   For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a

sequence comparison algorithm, test and reference sequences are input into a computer,

subsequence coordinates are designated, if necessary, and sequence algorithm program

parameters are designated. The sequence comparison algorithm then calculates the percent

sequence identity for the test sequence(s) relative to the reference sequence, based on the

5   designated program parameters.

[0146]   Optimal alignment of sequences for comparison can be conducted, *e.g.*, by

the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by

the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970),

by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA

10   85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT,

FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer

Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally* Ausubel *et

al.*, *infra*).

[0147]   One example of an algorithm that is suitable for determining percent

15   sequence identity and sequence similarity is the BLAST algorithm, which is described in

Altschul *et al.*, J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST

analyses is publicly available through the National Center for Biotechnology Information

(www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence

pairs (HSPs) by identifying short words of length W in the query sequence, which either

20   match or satisfy some positive-valued threshold score T when aligned with a word of the

same length in a database sequence. T is referred to as the neighborhood word score

threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for

initiating searches to find longer HSPs containing them. The word hits are then extended in

both directions along each sequence for as far as the cumulative alignment score can be

25   increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters

M (reward score for a pair of matching residues; always > 0) and N (penalty score for

mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to

calculate the cumulative score. Extension of the word hits in each direction are halted

when: the cumulative alignment score falls off by the quantity X from its maximum

30   achieved value; the cumulative score goes to zero or below, due to the accumulation of one

or more negative-scoring residue alignments; or the end of either sequence is reached. The

BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the

alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff &

5   Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

[0148]   In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an

10   indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

15   Mutagenesis and Other Molecular Biology Techniques
        [0149]   Polynucleotide and polypeptides of the invention and used in the invention can be manipulated using molecular biological techniques. General texts which describe molecular biological techniques include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA

20   (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2003) ("Ausubel")). These texts describe mutagenesis, the use of

25   vectors, promoters and many other relevant topics related to, e.g., the generation of genes that include selector codons for production of proteins that include selected amino acids (e.g., unnatural amino acids), glutamyl orthogonal tRNAs, glutamyl orthogonal synthetases, and pairs thereof.

        [0150]   Various types of mutagenesis are used in the invention, e.g., to mutate

30   glutamyl tRNA molecules, to produce libraries of glutamyl tRNAs, to produce libraries of glutamyl synthetases, to insert selector codons that encode a selected amino acid in a protein or polypeptide of interest. They include but are not limited to site-directed, random point

mutagenesis, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction, mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA or the like, or any combination thereof. Additional

5   suitable methods include point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, double-strand break repair, and the like. Mutagenesis, e.g., involving chimeric constructs, is also included in the invention. In one embodiment, mutagenesis can be guided by known information of the naturally occurring molecule or

10   altered or mutated naturally occurring molecule, e.g., sequence, sequence comparisons, physical properties, crystal structure or the like.

[0151]   Host cells are genetically engineered (e.g., transformed, transduced or transfected) with the polynucleotides of the invention or constructs which include a polynucleotide of the invention, e.g., a vector of the invention, which can be, for example, a

15   cloning vector or an expression vector. For example, the coding regions for the orthogonal tRNA, the orthogonal tRNA synthetase, and the protein to be derivatized are operably linked to gene expression control elements that are functional in the desired host cell. Typical vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the

20   particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in eukaryotes, or prokaryotes, or both (e.g., shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and/or integration in prokaryotes, eukaryotes, or preferably both. *See* Giliman & Smith,

25   Gene 8:81 (1979); Roberts, *et al.*, Nature, 328:731 (1987); Schneider, B., *et al.*, Protein Expr. Purif. 6435:10 (1995); Ausubel, Sambrook, Berger (*all supra*). The vector can be, for example, in the form of a plasmid, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide. The vectors are introduced into cells and/or microorganisms by standard methods including electroporation (From et al., Proc. Natl. Acad. Sci. USA 82,

30   5824 (1985), infection by viral vectors, high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., Nature 327, 70-73 (1987)), and/or the like.

[0152]   A catalogue of Bacteria and Bacteriophages useful for cloning is provided,

e.g., by the ATCC, e.g., The ATCC Catalogue of Bacteria and Bacteriophage (1996)

Ghema *et al.* (eds) published by the ATCC. Additional basic procedures for sequencing,

cloning and other aspects of molecular biology and underlying theoretical considerations are

5   also found in Sambrook (*supra*), Ausubel (*supra*), and in Watson *et al.* (1992) Recombinant

DNA Second Edition Scientific American Books, NY. In addition, essentially any nucleic

acid (and virtually any labeled nucleic acid, whether standard or non-standard) can be

custom or standard ordered from any of a variety of commercial sources, such as the

Midland Certified Reagent Company (Midland, TX mcrc.com), The Great American Gene

10   Company (Ramona, CA available on the World Wide Web at genco.com), ExpressGen Inc.

(Chicago, IL available on the World Wide Web at expressgen.com), Operon Technologies

Inc. (Alameda, CA) and many others.

[0153]   The engineered host cells can be cultured in conventional nutrient media

modified as appropriate for such activities as, for example, screening steps, activating

15   promoters or selecting transformants. These cells can optionally be cultured into transgenic

organisms. Other useful references, e.g. for cell isolation and culture (e.g., for subsequent

nucleic acid isolation) include Freshney (1994) Culture of Animal Cells, a Manual of Basic

Technique, third edition, Wiley- Liss, New York and the references cited therein; Payne *et*

al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New

20   York, NY; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture;

Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New

York) and Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC

Press, Boca Raton, FL.

PROTEINS AND POLYPEPTIDES OF INTEREST

25   [0154]   Methods of producing a protein in a cell with a selected amino acid at a

specified position are also a feature of the invention. For example, a method includes

growing, in an appropriate medium, the cell, where the cell comprises a nucleic acid that

comprises at least one selector codon and encodes a protein; and, providing the selected

amino acid, and incorporating the selected amino acid into the specified position in the

30   protein during translation of the nucleic acid with the at least one selector codon, thereby

producing the protein. The cell further comprises: an orthogonal glutamyl-tRNA (glutamyl-

O-tRNA) that functions in the cell and recognizes the selector codon; and, an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates the glutamyl-O-tRNA with the selected amino acid. Typically, the glutamyl O-tRNA comprises at least about, e.g., a 45%, a 50%, a 60%, a 75%, a 80%, or a 90% or more suppression efficiency in the presence of a cognate synthetase in response to the selector codon as compared to the glutamyl O-tRNA comprising or encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 67 (AE(GC) tRNA). A protein produced by this method is also a feature of the invention.

[0155]   The compositions of the invention and compositions made by the methods of the invention optionally are in a cell. The glutamyl O-tRNA/O-RS pairs or individual components of the invention can then be used in a host system's translation machinery, which results in a selected amino acid, e.g., unnatural amino acid, being incorporated into a protein. The International Application Number PCT/US2004/011786; and published Internation Application Number WO 2002/085923, *supra*, describe similar processes (e.g., with other orthogonal O-tRNA/O-RS pairs) and are incorporated herein by reference. For example, when an glutamyl O-tRNA/O-RS pair is introduced into a host, e.g., *Escherichia coli*, the pair leads to the in vivo incorporation of selected amino acid, such as an unnatural amino acid, e.g., a synthetic amino acid, such as derivative of a glutamate or glutamic acid amino acid, which can be exogenously added to the growth medium, into a protein, in response to a selector codon. Optionally, the compositions of the invention can be in an in vitro translation system, or in an in vivo system(s).

[0156]   Essentially any protein (or portion thereof) that includes a selected amino acid, e.g., an unnatural amino acid, (and any corresponding coding nucleic acid, e.g., which includes one or more selector codons) can be produced using the compositions and methods herein. No attempt is made to identify the hundreds of thousands of known proteins, any of which can be modified to include one or more unnatural amino acid, e.g., by tailoring any available mutation methods to include one or more appropriate selector codon in a relevant translation system. Common sequence repositories for known proteins include GenBank EMBL, DDBJ and the NCBI. Other repositories can easily be identified by searching the internet.

[0157]   Typically, the proteins are, e.g., at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 99% or more identical to any available

protein (e.g., a therapeutic protein, a diagnostic protein, an industrial enzyme, or portion thereof, and the like), and they comprise one or more selected amino acid.   Examples of therapeutic, diagnostic, and other proteins that can be modified to comprise one or more selected amino acid, e.g., an unnatural amino acid, can be found, but not limited to, those in

5   USSN 60/496,548, USSN 60/479,931, and USSN 60/463,869, *supra*; and, WO 2002/085923, *supra*.

[0158]   In certain embodiments, the protein or polypeptide of interest (or portion thereof) in the methods and/or compositions of the invention is encoded by a nucleic acid. Typically, the nucleic acid comprises at least one selector codon, at least two selector

10   codons, at least three selector codons, at least four selector codons, at least five selector codons, at least six selector codons, at least seven selector codons, at least eight selector codons, at least nine selector codons, ten or more selector codons.

[0159]   Genes coding for proteins or polypeptides of interest can be mutagenized using methods well-known to one of skill in the art and described herein under

15   "Mutagenesis and Other Molecular Biology Techniques" to include, e.g., one or more selector codon for the incorporation of a selected amino acid, e.g., an unnatural amino acid. For example, a nucleic acid for a protein of interest is mutagenized to include one or more selector codon, providing for the insertion of the one or more selected amino acids, e.g., unnatural amino acid. The invention includes any such variant, e.g., mutant, versions of any

20   protein, e.g., including at least one selected amino acid. Similarly, the invention also includes corresponding nucleic acids, i.e., any nucleic acid with one or more selector codon that encodes one or more selected amino acid.

[0160]   To make a protein that includes a selected amino acid, one can use host cells and organisms that are adapted for the in vivo incorporation of the selected amino acid via

25   orthogonal glutamyl tRNA/RS pairs. Host cells are genetically engineered (e.g., transformed, transduced or transfected) with one or more vectors that express the orthogonal glutamyl tRNA, the orthogonal glutamyl tRNA synthetase, and a vector that encodes the protein to be derivatized. Each of these components can be on the same vector, or each can be on a separate vector, or two components can be on one vector and the third component

30   on a second vector. The vector can be, for example, in the form of a plasmid, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide.

Defining Polypeptides by Immunoreactivity

[0161]   Because the polypeptides of the invention provide a variety of new

polypeptide sequences (e.g., comprising selected amino acids (e.g., unnatural amino acids)

in the case of proteins synthesized in the translation systems herein, or, e.g., in the case of

5   the novel synthetases, novel sequences of standard amino acids), the polypeptides also

provide new structural features which can be recognized, e.g., in immunological assays.

The generation of antisera, which specifically bind the polypeptides of the invention, as well

as the polypeptides which are bound by such antisera, are a feature of the invention. The

term "antibody," as used herein, includes, but is not limited to a polypeptide substantially

10   encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which

specifically bind and recognize an analyte (antigen). Examples include polyclonal,

monoclonal, chimeric, and single chain antibodies, and the like. Fragments of

immunoglobulins, including Fab fragments and fragments produced by an expression

library, including phage display, are also included in the term "antibody" as used herein.

15   *See, e.g.,* Paul, Fundamental Immunology, 4th Ed., 1999, Raven Press, New York, for

antibody structure and terminology.

[0162]   In order to produce antisera for use in an immunoassay, one or more of the

immunogenic polypeptides is produced and purified as described herein. For example,

recombinant protein can be produced in a recombinant cell. An inbred strain of mice (used

20   in this assay because results are more reproducible due to the virtual genetic identity of the

mice) is immunized with the immunogenic protein(s) in combination with a standard

adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (*see, e.g.,*

Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor

Publications, New York, for a standard description of antibody generation, immunoassay

25   formats and conditions that can be used to determine specific immunoreactivity.

[0163]   Additional details on proteins, antibodies, antisera, etc. can be found in

USSN 60/496,548, USSN 60/479,931, and USSN 60/463,869, *supra*; WO 2002/085923,

supra; patent application entitled "Glycoprotein synthesis" filed October 15, 2003, attorney

docket number 54-000610US; and patent application entitled "Protein Arrays," attorney

30   docket number P1001US00 filed on December 22, 2002.

USE OF GLUTAMYL O-tRNA AND GLUTAMYL O-RS AND GLUTAMYL O-tRNA/O-RS PAIRS

[0164]   The compositions of the invention and compositions made by the methods of the invention optionally are in a cell. The glutamyl O-tRNA/O-RS pairs or individual components of the invention can then be used in a host system's translation machinery, which results in a selected amino acid being incorporated into a protein. WO 2002/085923 by Schultz, et al., *supra*, describes related processes in general, and is incorporated herein by reference. For example, when an O-tRNA/O-RS pair is introduced into a host, e.g., *Escherichia coli*, the pair leads to the in vivo incorporation of a selected amino acid, which can be exogenously added to the growth medium, into a protein in response to a selector codon, e.g., an amber nonsense codon. Optionally, the compositions of the present invention can be in an in vitro translation system, or in an in vivo system(s). Proteins with the selected amino acid can be used as therapeutic proteins and can be used to facilitate studies on protein structure, interactions with other protein, and the like.

KITS

[0165]   Kits are also a feature of the invention. For example, a kit for producing a protein that comprises at least one selected amino acid, e.g., an unnatural amino acid, in a cell is provided, where the kit includes a container containing a polynucleotide sequence encoding an glutamyl O-tRNA, and/or an glutamyl O-tRNA, and/or a polynucleotide sequence encoding an glutamyl O-RS, and/or an glutamyl O-RS. In one embodiment, the kit further includes a selected amino acid. In another embodiment, the kit further comprises instructional materials for producing the protein.

EXAMPLES

[0166]   The following examples are offered to illustrate, but not to limit the claimed invention. One of skill will recognize a variety of non-critical parameters that may be altered without departing from the scope of the claimed invention.

EXAMPLE 1: AN ARCHAEBACTERIA-DERIVED GLUTAMYL-tRNA SYNTHETASE AND tRNA PAIR FOR UNNATURAL AMINO ACID MUTAGENESIS

[0167]   In this embodiment of the invention, the development of an orthogonal glutamic acid synthetase and tRNA pair is described. The tRNA is derived from the consensus sequence obtained from a multiple sequence alignment of, e.g., archaeal tRNA^{Glu}

sequences. The glutamyl-tRNA synthetase is from, e.g., the achaebacterium *Pyrococcus horikoshii*. The new orthogonal pair suppresses, e.g., amber, nonsense codons with an efficiency roughly comparable to that of the orthogonal tyrosine pair derived from *Methanococcus jannaschii*, which has been used to selectively incorporate a variety of unnatural amino acids into proteins in *E. coli*. Development of the glutamic acid orthogonal pair increases the potential diversity of unnatural amino acid structures that can be incorporated into proteins in *E. coli*, and can permit the *in vivo* incorporation of multiple distinct unnatural amino acids into the same protein.

[0168]   The ability to genetically incorporate unnatural amino acids into proteins offers new opportunities to investigate protein structure and function *in vitro* and *in vivo*, as well as produce large quantities of proteins with novel properties. This approach involves the directed evolution of the amino acid specificity of an aminoacyl-tRNA synthetase and tRNA pair that is orthogonal in, e.g., *E. coli*. *See, e.g.,* 1. Liu,D.R. and Schultz,P.G. (1999) *Progress toward the evolution of an organism with an expanded genetic code*. Proc. Natl. Acad. Sci. USA, 96:4780-4785; Wang,L., Brock,A., Herberich,B. and Schultz,P.G. (2001) *Expanding the genetic code of Escherichia coli*. Science, 292:498-500; and, Santoro,S.W., Wang,L., Herberich,B., King,D.S. and Schultz,P.G. (2002) *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity*. Nat. Biotechnol. 20:1044-1048. Such orthogonal pairs satisfy several criteria: the tRNA is not a substrate for any of the endogenous *E. coli* synthetases but functions efficiently in protein translation; the orthogonal synthetase efficiently aminoacylates the orthogonal tRNA, the anticodon of which has been modified to recognize a selector codon, e.g., an amber (UAG) or opal (UGA) stop codon, a four-base codon, etc.; and the synthetase does not aminoacylate any of the endogenous *E. coli* tRNAs. To date, several orthogonal pairs have been developed for use in *E. coli*, including glutamine (*see, e.g.,* Liu,D.R. and Schultz,P.G. (1999) *Progress toward the evolution of an organism with an expanded genetic code*. Proc. Natl. Acad. Sci. USA, 96:4780-4785), aspartic acid *(see, e.g.,* Pastrnak,M., Magliery,T.J. and Schultz,P.G. (2000) *A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code*. Helvetica Chimica Acta, 83:2277-2286) and tyrosine (*see, e.g.,* Ohno,S., Yokogawa,T., Fujii,I., Asahara,H., Inokuchi,H. and Nishikawa,K. (1998) *Co-expression of yeast amber suppressor tRNATyr and tyrosyl-tRNA synthetase in Escherichia coli: possibility to expand the genetic code*. J. Biochem. (Tokyo),

124:1065-1068; and, Kowal,A.K., Kohrer,C. and RajBhandary,U.L. (2001) *Twenty-first*
aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific
incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. Proc.
Natl. Acad. Sci. USA, 98:2268-2273) pairs from *Saccharomyces cerevisiae*, a tyrosine pair
5   from *Methanococcus jannaschii* (*see, e.g.*, Wang,L., Magliery,T.J., Liu,D.R. and
Schultz,P.G. (2000) *A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for*
the in vivo incorporation of unnatural amino acids into proteins. J. Am. Chem. Soc.,
122:5010-5011), and a leucine pair from *Methanobacterium thermoautotrophicum* (*see,*
e.g., Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-*
10   *tRNA and synthetase pair for four-base, amber, and opal suppression*, Biochemistry, Aug
19; 42(32):9598-9608). In addition, glutamine (*see, e.g.*, Kowal,A.K., Kohrer,C. and
RajBhandary,U.L. (2001) *Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs*
for possible use in site-specific incorporation of amino acid analogues into proteins in
eukaryotes and in eubacteria. Proc. Natl. Acad. Sci. USA, 98:2268-2273) and tyrosine (*see,*
15   *e.g.*, Edwards and Schimmel (1990) *A bacterial amber suppressor in Saccharomyces*
cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase. Mol. Cell.
Biol., 10:1633-1641) synthetase-tRNA pairs from *E. coli* have been developed for use as
orthogonal pairs in *S. cerevisiae*.

[0169]   The orthogonal tyrosine pair from *M. jannaschii* has been successfully used
20   to incorporate unnatural amino acids into proteins in *E. coli*. Many if not all of the unnatural
amino acids that have been incorporated thus far using variants of the tyrosyl-tRNA
synthetase are aryl derivatives, including *p*-azido- (*see, e.g.*, Chin,J.W., Santoro,S.W.,
Martin,A.B., King,D.S., Wang,L. and Schultz,P.G. (2002) *Addition of p-azido-L-*
phenylalanine to the genetic code of Escherichia coli. J. Am. Chem. Soc., 124:9026-9027),
25   *p*-benzoyl- (*see, e.g.*, Chin,J.W., Martin,A.B., King,D.S., Wang,L. and Schultz,P.G. (2002)
Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. Proc.
Natl. Acad. Sci. USA, 99:11020-11024), *p*-amino- (*see, e.g.*, Santoro,S.W., Wang,L.,
Herberich,B., King,D.S. and Schultz,P.G. (2002) *An efficient system for the evolution of*
aminoacyl-tRNA synthetase specificity. Nat. Biotechnol. 20:1044-1048), *p*-isopropyl- (*see,*
30   *e.g.*, Santoro,S.W., Wang,L., Herberich,B., King,D.S. and Schultz,P.G. (2002) *An efficient*
system for the evolution of aminoacyl-tRNA synthetase specificity. Nat. Biotechnol.
20:1044-1048), *m*-acetyl-(*see, e.g.*, Wang,L., Zhang,Z., Brock,A. and Schultz,P.G. (2003)

Addition of the keto functional group to the genetic code of Escherichia coli. Proc. Natl. Acad. Sci. USA, 100:56-61), and *p*-acetyl-phenylalanine (*see, e.g.,* Wang,L., Zhang,Z., Brock,A. and Schultz,P.G. (2003) *Addition of the keto functional group to the genetic code of Escherichia coli*. Proc. Natl. Acad. Sci. USA, 100:56-61); *O*-methyl- (*see, e.g.,* Wang,L.,

5   Brock,A., Herberich,B. and Schultz,P.G. (2001) *Expanding the genetic code of Escherichia coli*. Science, 292:498-500) and *O*-allyl-tyrosine (Santoro,S.W., Wang,L., Herberich,B., King,D.S. and Schultz,P.G. (2002) *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity*. Nat. Biotechnol. 20:1044-1048; and, Zhang,Z., Wang,L., Brock,A. and Schultz,P.G. (2002) *The selective incorporation of alkenes into proteins in Escherichia*

10   *coli*. Angew. Chem. Int. Ed. Engl., 41:2840-2842); 3-(2-naphthyl)alanine (*see, e.g.,*Wang,L., Brock,A. and Schultz,P.G. (2002) *Adding L-3-(2-Naphthyl)alanine to the genetic code of E. coli*. J. Am. Chem. Soc., 124:1836-1837), etc. In order to increase the structural and chemical diversity of unnatural amino acids that can be incorporated into proteins expressed by, e.g., *E. coli*, additional active orthogonal pairs would be useful.

15   **[0170]**   Archaea appear to be an especially good source of orthogonal pairs for *E. coli*. Previous studies revealed that several archaeal tRNAs are not recognized by *E. coli* aminoacyl-tRNA synthetases (*see, e.g.,* Kwok,Y. and Wong,J.T. (1980) *Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes*. Can. J. Biochem., 58:213-218). Two

20   such synthetase-tRNA pairs derived from archaea, for tyrosine (*see e.g.,* Wang,L., Magliery,T.J., Liu,D.R. and Schultz,P.G. (2000) *A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins*. J. Am. Chem. Soc., 122:5010-5011) and leucine (*see, e.g.,* Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase*

25   *pair for four-base, amber, and opal suppression*, Biochemistry, Aug 19; 42(32):9598-9608), were developed for use in *E. coli*. Both exhibited high levels of orthogonality and amber suppression efficiency. Moreover, these archaeal synthetases and tRNAs are efficiently produced and processed in *E. coli*. Finally, the recent proliferation of genomic sequence data for archaeal species and the commercial availability of their genomic DNA

30   have facilitated the cloning of synthetases from these organisms as well as the design of compatible tRNAs.

[0171]   Archaeal glutamyl-tRNA synthetases and tRNAs are attractive starting points for the generation of additional orthogonal pairs for several reasons. Archaeal glutamyl-tRNA synthetases are expected to tolerate changes in the anticodon loop of their substrate tRNAs since the enzyme naturally accommodates four different tRNA anticodons:

5   UUC, CUC, UUG, and CUG. This is due to the fact that in archaeal species, glutamyl-tRNA synthetases must recognize and acylate both $tRNA^{Glu}$ and $tRNA^{Gln}$ (see, e.g., Tumbula,D.L., Becker,H.D., Chang,W.Z. and Soll,D. (2000) Domain-specific recruitment of amide amino acids for protein synthesis. Nature, 407:106-110). It follows that an archaeal glutamyl-tRNA synthetase should tolerate variance within at least two positions of

10   the anticodon of its tRNA substrate, which should permit efficient aminoacylation of a tRNA bearing an amber-suppressing CUA anticodon. In addition, in vitro aminoacylation data (see, e.g., Kwok,Y. and Wong,J.T. (1980) Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. Can. J. Biochem., 58:213-218) predict that archaeal

15   glutamic acid and glutamine tRNAs should be orthogonal in E. coli. Finally, the three dimensional structure of the glutamyl-tRNA synthetase from Thermus thermophilus is known (see, e.g., Nureki,O., Vassylyev,D.G., Katayanagi,K., Shimizu,T., Sekine,S., Kigawa,T., Miyazawa,T., Yokoyama,S. and Morikawa,K. (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. Science, 31:1958-1965; and,

20   Sekine,S., Nureki,O., Shimada,A., Vassylyev,D.G. and Yokoyama,S. (2001) Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase. Nat. Struct. Biol., 8:203-206), which should facilitate the design of libraries of synthetase variants for use in directed evolution experiments.

Materials and Methods

25   [0172]   Bacterial strains, plasmids, and reagents: E. coli strains DH5α and DH10B were purchased from Stratagene (La Jolla, CA) and Invitrogen Life Sciences (Carlsbad, CA), respectively. Archaeoglobus fulgidus (Af), Aeropyrum pernix (Ap), Methanococcus jannaschii (Mj), Methanosarcina mazei Goe1 (Mm), Methanobacterium thermoautotrophicum (Mt), Pyrococcus horikoshii (Ph), and Sulfolobus solfataricus (Ss)

30   genomic DNA was obtained from ATCC (Manassas, VA). E. coli genomic DNA was prepared from strain DH5α. Plasmid pKQ, which contains a ColE1 origin of replication, a kanamycin resistance marker, and cloning sites for an aminoacyl-tRNA synthetase gene, has

been described previously (*see, e.g.,* Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression,* Biochemistry, Aug 19; 42(32):9598-9608). Plasmid pACKO-A184TAG, which contains a p15A origin of replication, a chloramphenicol resistance marker, a β-

5   lactamase ampicillin resistance gene containing an amber stop codon at position Ala184, and tRNA cloning sites, has been described previously (*see, e.g.,* Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression,* Biochemistry, Aug 19; 42(32):9598-9608). Plasmid pArgU218, which contains a p15A origin of replication and encodes a kanamycin

10   resistance marker and the isoacceptor of tRNA^{Arg} corresponding to the codons AGA and AGG, has been described previously (*see, e.g.,* Bullard,J.M., Cai,Y.C. and Spremulli,L.L. (2000) *Expression and characterization of the human mitochondrial leucyl-tRNA synthetase.* Biochim. Biophys. Acta, 1490:245-258). Plasmid pBAD-Myc/HisA was purchased from Invitrogen. PCR reactions were carried out using the Expand Kit (Roche,

15   Indianapolis, IN). Whole *E. coli* tRNA was purchased from Roche. Whole halobacterial tRNA was isolated from cultures of *Halobacterium sp. NRC-1* using the RNA/DNA Extraction Kit (Qiagen). Overlap extensions were carried out using Taq polymerase (Stratagene). Restriction enzymes were purchased from New England Biolabs (Beverly, MA). DNA was purified using miniprep, gel purification, or PCR purification kits (Qiagen).

20   Oligonucleotides were prepared by Qiagen Operon (Alameda, CA). Sequences of oligodeoxyribonucleotides used in this example:

pKQ-AfERS.N: d(GGAGAGAAGACTACATGAAAGAAGTCATAATGAAATAC)

pKQ/pBAD-AfERS.C: d(CGTTAGAAGACTGAATTCGGATTTTGGTCACGGGTGCG)

pKQ-ApERS.N: d(GCTATCACACCATGGCGATGCTCCTAGGGGACCATC)

25   pKQ-ApERS.C: d(GATCAGTCAGAATTCTTACTTGTGTGTGTATATCACCCTG)

pKQ-MjERS.N: d(GGAGACGTCTCACATGGAAGAGAAGATATTGCC)

pKQ-MjERS.C: d(CGTTACGTCTCGAATTTTATCTATGTGCATAGCAAC)

pKQ-MmERS.N: d(GGAGACGTCTCACATGACCTTAAGTCCTGAAGA)

pKQ/pBAD-MmERS.C: d(CGTTACGTCTCGAATTTTATTTATGAGCAAAGTACGCCACGAC)

30   pKQ-MtERS.N: d(GGAGACGTCTCTCATGGTGCCAGTGGAGGACCT)

pKQ/pBAD-MtERS.C: d(CGTTACGTCTCTAATTTTATTTGTGGGCATAGTAGAAG)

pKQ-PhERS.N: d(GGAAAGGTCTCTCATGGATGTGGAAAAGATAGC)

pKQ/pBAD-PhERS.C: d(GAGTAGGTCTCTAATTTTACTTGTGAGCATATATTG)

pKQ-SsERS.N: d(GCTACTGCACCATGGAATTAAATGAATTACGAGAAC)

pKQ-SsERS.C: d(GACTCAGTAGAATTCATTAGTCGTGTGAGAATACAACA)

5   AEGU-tRNA.F: d(CGGAATTCGCCCCGGTGGTGTAGCCCGGCCAAGCATGCGGGCCTCTAA)

AEGU-RNA.R: d(AAACTGCAGTGGTGCCCCGGCCGGGATTTGAACCCGGGTCGCGGGCTTTAG

        AGGCCCGCATGCTTGG)

AQGU-RNA.F: d(CGGAATTCAGCCCCGTGGTGTAGCGGCCAAGCATGCGGGGCTCTAAACCCC

        GCGACCG)

10 AQGU-RNA.R: d(AAACTGCAGTGGTAGCCCCGCGGGGATTCGAACCCCGGTCGCGGGGTTTAG

        AGCC)

AQGC-RNA.F: d(CGGAATTCAGCCCCGTGGTGTAGCGGCCAAGCACGCGGGGCTCTAAACCCC

        GCGACCG)

AQGC-RNA.R: d(AAACTGCAGTGGTAGCCCCGCGGGGATTCGAACCCCGGTCGCGGGGTTTAG

15   AGCC)

AENN-RNA.F: d(CGGAATTCGCCCCGGTGNTGTAGCCCGGCCAAGCANGCGGGCCTCTAA)

pAC-tRNA.F: d(GGACAGCGGTGCGGACTG)

pAC-tRNA.R: d(GTGCTCAACGGCCTCAAC)

pBAD-EcERS.N: d(AGGCTTTTAACCATGGCAAAAATCAAAACTCGCTTCGC)

20 pBAD-EcERS.C: d(ACGTGAGAGGAATTCTGCTGATTTTCGCGTTCA)

pBAD-AfERS.C: d(CGTTAGAAGACTGAATTGGGGTGAGTGTACACCGCGACCAGTTCCTCG)

pBAD-MmERS.C: d(CGTTACGTCTCGAATTCTTATGAGCAAAGTACGCCACGAC)

pBAD-MtERS.C: d(CGTTACGTCTCTAATTTTTGTGGGCATAGTAGAAGACC)

pBAD-PhERS.C: d(GAGTAGGTCTCTAATTCTTGTGAGCATATATTGCGATAA)

25 **[0173]**   *Cloning of aminoacyl-tRNA synthetase genes*: The open reading frames of
glutamyl-tRNA synthetase from *A. fulgidus* (AfERS), *A. pernix* (ApERS), *M. jannaschii*
(MjERS), *M. mazei* (MmERS), *M. thermoautotrophicum* (MtERS), *P. horikoshii* (PhERS),
and *S. solfataricus* (SsERS) were amplified by PCR using genomic DNA as template and
the appropriate oligonucleotide primers (see above) to yield products of ~1.7 kilo-base pairs
30 (kbp). In these experiments, the amplification of the MjERS gene was unsuccessful. The

AfERS PCR product was digested with BbsI, the ApERS and SsERS products with EcoRI and NcoI, the MmERS and MtERS products with BsmBI, and the PhERS product with BsaI to generate EcoRI / NcoI compatible cohesive ends for all products. Construction of plasmids pKQ-AfERS, pKQ-ApERS, pKQ-MjERS, pKQ-MmERS, pKQ-MtERS, pKQ-

5   PhERS, and pKQ-SsERS was accomplished by insertion of restriction-digested, gel-purified glutamyl-tRNA synthetase gene PCR fragments into plasmid pKQ. Plasmid constructs were confirmed by restriction mapping and sequencing of individual clones.

[0174]   *Design and cloning of tRNAs*: Consensus archaeal tRNA genes were identified by alignment using the program *pileup* from Genetics Computer Group, Inc.

10   (GCG). Genes for tRNAs were constructed by overlap extension of the appropriate forward and reverse primers (see above). *EcoRI/PstI*-digested extension fragments were inserted into a 4.2 kbp fragment of plasmid pACKO-A184TAG to construct plasmids pAC-AE(GU), pAC-AQ(GU), and pAC-AQ(GC). The AE(NN) tRNA library was constructed by PCR amplification using primers AE(NN)-RNA.F and pAC-tRNA.R with plasmid pAC-AE(GU)

15   as template. The pAC-AE(NN) plasmid library was constructed by insertion of the *EcoRI/PstI*-digested AE(NN) tRNA library PCR fragment into the tRNA cloning site of pACKO-A184TAG. Plasmid constructs were confirmed by sequence analysis.

Nomenclature of tRNAs:

AQ(GU): A-archaeal; Q-glutaminyl; GU-base pair at postion 10-26

20   AQ(GC): A-archaeal; Q-glutaminyl; GC-base pair at postion 10-26

AE(GU): A-archaeal; E-glutamyl; GU-base pair at postion 10-28

AE(GC): A-archaeal; E-glutamyl; GC-base pair at postion 10-28

AE(NN): A-archaeal; E-glutamyl; NN-base pair at postion 10-28

[0175]   *Ampicillin IC_{50} value measurements*: To examine the orthogonality of

25   tRNAs, 10 μL of *E. coli* DH10B cells (1 cfu/μL) carrying a tRNA-expressing plasmid were plated on a series of LB agar plates containing chloramphenicol and varying concentrations of ampicillin. To measure the IC_{50} values for cells carrying both tRNA- and synthetase-expressing plasmids, cells were plated on LB agar plates containing chloramphenicol, kanamycin, and varying concentrations of ampicillin. In the initial series, cells were plated

30   on five ampicillin concentrations varying from 0 to 1000 μg/mL and incubated overnight at 37°C. In subsequent series, cells were plated on a narrower ampicillin concentration range

surrounding the value initial ampicillin IC_{50} value estimate. Typically, 10 ampicillin concentrations spanning a range ± 45% of the initial IC_{50} estimate were tested in 10% increments. Measurements were made in duplicate or triplicate for each cell strain.

[0176]   *Optimization of the AEGU tRNA*: *E. coli* DH10B cells carrying plasmid pKQ-MmERS were transformed with the pAC-AE(NN) plasmid library to yield ~10^6 independent transformants. Cells were plated (~1000 cfu/plate) on a series of ten LB agar plates containing chloramphenicol, kanamycin, and 0–1000 µg/mL ampicillin. Ten individual colonies from the plate containing 500 µg/mL ampicillin were picked and their pAC-AE(NN) plasmids amplified, miniprepped, and sequenced.

[0177]   *Synthetase overexpression*: Aminoacyl-tRNA synthetase genes were PCR-amplified using the appropriate N- and C-terminal pBAD construct primers (*see* Bacterial strains, plasmids and reagents) and the appropriate synthetase pKQ construct templates. The AfERS PCR product was digested with *Bbs*I, the EcERS product with *Eco*RI and *Nco*I, the MmERS and MtERS products with *Bsm*BI, and the PhERS product with *Bsa*I to generate *Eco*RI / *Nco*I compatible cohesive ends for all products. Construction of the synthetase pBAD plasmids was accomplished by insertion of restriction-digested, gel-purified PCR fragments into the *Eco*RI and *Nco*I cloning sites of plasmid pKQ. Plasmid constructs were confirmed by restriction mapping and sequencing of individual clones. The synthetase pBAD plasmids were transferred to DH10B cells carrying plasmid pArgU218. Individual colonies were used to inoculate 500-mL cultures of 2YT containing ampicillin and kanamycin. Protein was isolated according to the QIAexpressionist (Qiagen) protocol under native conditions and dialyzed overnight against 20 mM Tris-HCl (pH 8), 200 mM NaCl, and 1 mM DTT. Glycerol was added to protein samples to a final concentration of 50%. Glycerol-containing samples were stored at -20°C. Protein concentrations were determined using sodium-dodecyl sulfate polyacrylamide gel electrophoresis by comparison to a BSA standard.

[0178]   *In vitro aminoacylation assays*: Reactions were carried out as described previously (*see, e.g.,* Hoben,P. and Soll,D. (1985) *Glutaminyl-tRNA synthetase of Escherichia coli*. Methods Enzymol., 113:55-59). Each 20-µL reaction contained 50 mM Tris-HCl (pH 7.5), 30 mM KCl, 20 mM $MgCl_2$, 0.1 mg/mL BSA, 5 mM glutathione, 2.5 mM ATP, 100 nM l-[G-^{3}H]Glutamic acid (Amersham), 750 nM aminoacyl-tRNA

synthetase, and 40 µM whole halobacterial tRNA. Reactions were allowed to proceed for 15 minutes at 37°C.

Results and Discussion

[0179]   *Development of an orthogonal tRNA*: An orthogonal tRNA derived from the
consensus sequence of archaeal leucine tRNAs can serve as an efficient substrate for
archaeal leucyl-tRNA synthetases. *See, e.g.,* the generation of an orthogonal tRNA^{Leu} (*see
e.g.,* Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-
tRNA and synthetase pair for four-base, amber, and opal suppression,* Biochemistry, Aug
19;42(32):9598-9608). The design of an orthogonal tRNA using this strategy involves:
alignment of a group of related tRNA sequences to determine the consensus sequence;
analysis of the secondary structure of the consensus-derived tRNA to identify and correct
potentially deleterious non-canonical base pairs; and conversion of the anticodon of the
consensus tRNA to CUA to allow recognition and suppression of an amber stop codon
within the target mRNA. The same strategy was applied to construct a cognate orthogonal
tRNA for archaeal glutamyl-tRNA synthetases.

[0180]   In archaebacteria, a single enzyme, glutamyl-tRNA synthetase,
aminoacylates both tRNA^{Glu} and tRNA^{Gln}. The product of the latter reaction, glu-tRNA^{Gln},
is subsequently converted to gln-tRNA^{Gln} by transamidation (*see, e.g.,* Tumbula,D.L.,
Becker,H.D., Chang,W.Z. and Soll,D. (2000) *Domain-specific recruitment of amide amino
acids for protein synthesis.* Nature, 407:106-110). Previous reports suggest that both
archaeal tRNA^{Glu} and archaeal tRNA^{Gln} are not substrates for endogenous *E. coli*
synthetases (*see, e.g.,* Kwok,Y. and Wong,J.T. (1980) *Evolutionary relationship between
Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA
synthetases as phylogenetic probes.* Can. J. Biochem., 58:213-218). Thus, both tRNAs were
viewed as attractive starting points for the development of an orthogonal tRNA using the
consensus sequence design strategy. The available archaeal tRNA^{Glu} and tRNA^{Gln}
sequences were separately aligned, revealing a high degree of homology within each family
(Tables 1 and 2). The consensus sequence of each family was derived by identification of
the consensus base at each position within the primary sequence. The consensus tRNA^{Glu}
sequence contains three positions outside of the anticodon (positions 3, 25, and 72) at which
two bases predominate with approximately equal frequency (Table 1). The consensus
tRNA^{Gln} sequence contains five positions outside the anticodon (positions 16, 32, 40, 51 and

63) at which two bases predominate with approximately equal frequency and two positions (positions 6 and 67) at which one base predominates but does not represent a majority at its position (Table 2).

[0181]   The secondary structures of the consensus archaeal tRNA^{Glu} and tRNA^{Gln} sequences (**Figure 1, Panel A**) were used to design three candidate orthogonal tRNAs (**Figure 1, Panel B**). The anticodon loops contained the sequence CUA for suppression of an amber stop codon (UAG). The nucleotide to the 3′ side of each anticodon was converted to A to improve suppression efficiency in *E. coli* (*see, e.g.,* Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression*, Biochemistry, Aug 19;42(32):9598-9608). Consensus tRNA^{Glu} and tRNA^{Gln} structures contain a conserved non-canonical G-U base pair at positions 10-28 and 10-26, respectively (**Figure 1, Panel A**). Although it has been demonstrated that alteration of non-canonical base pairs within orthogonal tRNAs can improve amber stop codon suppression efficiency in *E. coli* (*see, e.g.,* Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression*, Biochemistry, Aug 19;42(32):9598-9608), the conserved nature of this pair suggested that it might constitute an important identity element for recognition by archaeal glutamyl-tRNA synthetases. Therefore, in designing the tRNA^{Glu} orthogonal tRNA, AE(GU), the 10-28 G-U base pair was preserved (**Figure 1, Panel B**). In designing the tRNA^{Gln} orthogonal tRNAs, AQ(GU) and AQ(GC), the 10-26 pair was either preserved or changed to a G-C pair, respectively.

TABLE 1. ARCHAEAL GLUTAMIC ACID AND CONSENSUS tRNAs

tRNA (anticodon)	Sequence[1]
Aeropyrum pernix (CUC)	GCCGCGGUAGUAUAGCCCGGCCCAGUAUGCGGGCCUCUCGAGCCCGU GACCCGGGUUCAAAUCCCGGCCGCGGCACCA
Aeropyrum pernix (UUC)	GCCGCGGUAGUAUAGCCCGGCCCAGUAUGCGGGCCUUUCGAGCCCGU GACCCGGGUUCAAAUCCCGGCCGCGGCACCA
Archaeoglobus fulgidus (CUC)	GCUCCGGUGGUGUAGCCCGGCCAAUCAUACGGGACUCUCGAUCCCGU GACCCGGGUUCAAAUCCCGGCCGGAGCACCA
Archaeoglobus fulgidus (UUC)	GCUCCGGUGGUGUAGCCCGGCCAAUCAUUCCGGCCUUUCGAGCCGGC GACCCGGGUUCAAAUCCCGGCCGGAGCACCA
Halobacterium sp. (CUC)	GCUCCGUUGGUGUAGUCCGGCCAAUCAUCUUGCCCUCUCACGGCAAG GACUAGGGUUCAAAUCCCUGACGGAGCACCA
Halobacterium sp. (UUC)	GCUCGGUUGGUGUAGUCCGGCCAAUCAUGUUGGCCUUUCGAGCCGA CGACCAGGGUUCAAAUCCCUGACCGAGCACCA
Methanobacterium thermautotrophicus (UUC)	GCUCCGGUAGUGUAGUCCGGCCAAUCAUUUCGGCCUUUCGAGCCGAA GACUCGGGUUCAAAUCCCGGCCGGAGCACCA

-67-

tRNA (anticodon)	Sequence[1]
Methanococcus jannaschii (UUC)	GCUCCGGUGGUGUAGUCCGGCCAAUCAUGCGGGCCUUUCGAGCCCGC GAGCCCGGGUUCAAAUCCCGGCCGGAGCACCA
Methanococcus maripaludis (UUC)	GCUCCAGUGGUGUAGUCCGGCCAAUCAUCCGGCCCUUUCGAGGCCGG GACUCGGGUUCAAAUCCCGGCUGGAGCACCA
Methanosarcina mazei (CUC)	GCUCCGAUAGUGUAGCUCGGCCAAUCAUUCAGGACUCUCACUCCUGC GACUGGGGUUCAAAUCCCCAUCGGAGCACCA
Methanosarcina mazei (UUC)	GCUCCGGUAGUGUAGUCCGGCCAAUCAUUCCGGCCUUUCGAGCCGAA GACUCGGGUUCGAAUCCCGGCCGGAGCACCA
Pyrobaculum aerophilum (CUC)	CAGCCGGUAGUCUAGCCCGGACAAGGAUGCGGGCCUCUCGAGCCCGU GACCCGGGUUCAAAUCCCGGCCGGGGCACCA
Pyrococcus abyssi (CUC)	GCCCCGGUGGUGUAGCCCGGUCAAUCAUGCGGGACUCUCGAUCCCGC GACCCGGGUUCAAAUCCCGGCCGGGGCACCA
Pyrococcus abyssi (UUC)	GCCCCGGUGGUGUAGCCCGGUCAAACAUGCGGGCCUUUCGAGCCCGC GCCCCGGGUUCAAAUCCCGGCCGGGGCACCA
Pyrococcus furiosus (CUC)	GCCCCGGUGGUGUAGCCCGGUCAAUCAUGCGGGACUCUCGAUCCCGC GACCCGGGUUCAAAUCCCGGCCGGGGCACCA
Pyrococcus furiosus (UUC)	GCCCCGGUGGUGUAGCCCGGUCAAACAUGCGGGCCUUUCGAGCCCGC GCCCCGGGUUCAAAUCCCGGCCGGGGCACCA
Pyrococcus horikoshii (CUC)	GCCCCGGUGGUGUAGCCCGGUCAAUCAUGCGGGACUCUCGAUCCCGC GACCCGGGUUCAAAUCCCGGCCGGGGCACCA
Pyrococcus horikoshii (UUC)	GCCCCGGUGGUGUAGCCCGGCCAAACAUGCGGGCCUUUCGAGCCCGC GCCCCGGGUUCAAAUCCCGGCCGGGGCACCA
Sulfolobus solfataricus (CUC)	GCCGCGGUGGUUUAGCCCGGUC- AGAAUGCGGGCCUCUCGAGCCCGUGACCCGGGUUCAAAUCCCGGCC GCGGCACCA
Sulfolobus solfataricus (UUC)	GCCGCGGUGGUUUAGCCCGGUC- AGAAUGCGGGCCUUUCGAGCCCGUGACCCGGGUUCAAAUCCCGGCCG CGGCACCA
Sulfolobus tokodaii (CUC)	GCCGCGGUAGUAUAGCCCGGUC- AGUAUGCGGGCCUCUCAAGCCCGUGGCCCGGGUUCAAAUCCCGGCCG CGGCACCA
Sulfolobus tokodaii (UUC)	GCCGCGGUAGUAUAGCCCGGUCAAGCACGCGGGCCUUUCGAGCCCGU GACCCGGGUUCAAAUCCCGGCCGCGGCACCA
Thermoplasma acidophilum (CUC)	GCUCCGGUGGUGUAGUCCGGCCAAGCAUUAGGGACUCUCAAUCCCCC GACUCGGGUCCAAAUCCCGACCGGAGCACCA
Thermoplasma acidophilum (UUC)	GCUCCGGUGGUGUAGUCCGGACAAGCAUAUCGGCCUUUCGAGCCGA CGACCUGGGUCCAAAUCCCGGCCGGAGCACCA
Thermoplasma volcanium (CUC)	GCUCCGGUGGUGUAGUCCGGCCAAGCAUUAGGGACUCUCAAUCCCCC GACUCGGGUCCAAAUCCCGACCGGAGCACCA
Thermoplasma volcanium (UUC)	GCUCCGGUGGUGUAGUCCGGACAAGCAUAUCGGCCUUUCGAGCCGA CGACCUGGGUCCAAAUCCCGGCCGGAGCACCA
Consensus (YUC)[2]	GCYCCGGUGGUGUAGCCCGGCCAAKCAUGCGGGCCUYUCGAGCCCG CGACCCGGGUUCAAAUCCCGGCCGGRGCACCA
AE(GU) (CUA)[3]	GCcCCGGUGGUGUAGCCCGGCCAAgCAUGCGGGCCUcUaaAGCCCGCGA CCCGGGUUCAAAUCCCGGCCGGgGCACCA

[1] Bases indicated by underlining deviate from the identity of their consensus bases. Bases highlighted in black correspond to positions at which two different bases predominate with approximately equal frequency. Anticodon positions are boxed.

[2] Bases that are in boldface correspond to positions involved in non-canonical base pairs (Figure 1, Panel A). Y corresponds to bases C or U; R corresponds to bases A or G; S corresponds to bases C or G.

[3] Bases in lowercase correspond to positions that were varied from the consensus sequence in the design of a prospective orthogonal tRNA (Figure 1, Panel B).

TABLE 2. ARCHAEAL GLUTAMINE AND CONSENSUS tRNAs

tRNA (anticodon)	Sequence[1]
Aeropyrum pernix (CUG)	AGCCGGGUCGUCUAGCGGCCCAGGAUGCGGGGCUCUGGCCCC CGUGACCGGGGUUCGAAUCCCCGCCCGGCUACCA
Aeropyrum pernix (UUG)	AGCCGGGUCGUCUAGCGGCCCAGGAUGCGGGGUUUUGGCCCC CGUGACCCGGGGUUCGAAUCCCGGCCCGGCUACCA
Archaeoglobus fulgidus (CUG)	AGCCCCGUGGGGUAGCGGUCAAUCCUGCCGGACUCUGGAUCC GGCGACGCCGGGUUCGAAUCCCGGCCGGGGCUACCA
Archaeoglobus fulgidus (UUG)	AGCCCCGUGGGUGUAGCGGUCAAUCAUGCGGGCCUUUUGGAGCC CGCGACCGGGGUUCGAAUCCCCGCGGGGCUACCA
Halobacterium sp. (CUG)	AGUCCCGUAGGGUAGUGGCCAAUCCUGAAGCCUUCUGGGGGC UUCGACGGAAGUUCGAAUCUUCCCGGGACUACCA
Halobacterium sp. (UUG)	AGUCCCGUGGUGUAGUGGCCAAUCAUAUGGGCCUUUUGGAGCC CACGACGGCGGGUUCGAAUCCGCCCCGGGACUACCA
Methanobacterium thermautotrophicus (CUG)	AGUCCCGUGGGGUAAUGGCA- AUCCUGAUGGACUCUGGAUCCAUCGAUAGCGGUUCGACUCCG CUCGGGACUACCA
Methanobacterium thermautotrophicus (UUG)	AGUCCCGUGGGGUAGUGGUA- AUCCUGCUGGGCUUUGGACCCGGCGACAGCGGUUCGACUCCG CUCGGGACUACCA
Methanococcus jannaschii (UUG)	AGCCCGGUGGUGUAGUGGCCUAUCAUCCGGGGCUUUUGGACCC CGGGACCGCGGGUUCGAAUCCGCGCCGGGCUACCA
Methanococcus maripaludis (UUG)	AGCCCAGUAGUGUAGUGGCCAAUCAUCCGUGCCUUUUGGAGCA UGGGACCGCGGGUUCGAAUCCGCGCUGGGCUACCA
Methanopyrus kandleri (UUG)	AGGGCCGUGGGGUAGCGGUCUAUCCUGCGGGGCUUUUGGACCC CGCGACCCCGGGUUCAAAUCCCGGCGGCCCUACCA
Methanosarcina mazei (CUG)	AGCCCGGUAGUGUAGUGGUCAAUCAUGCGGGACUCUGGACCC UGCAACCUCGGGUUCGAAUCCGUGCCGGGCUACCA
Methanosarcina mazei (UUG)	AGUCCUGUAGGGUAGUGGUCAAUCCUUCGGGCCUUUUGGAGCC CGGGACAGCGGUUCGAAUCCGCUCAGGACUACCA
Pyrobaculum aerophilum (CUG)	AGCCCGGUCGUCUAGCGGCC- AGGAUGCGGGGCUCUGGACCCCGUGGCCCGGGUUCGAAUCCC GGCCGGGCUACCA
Pyrobaculum aerophilum (UUG)	AGCCCGGUCGUCUAGCGGCCAAGGAUGCGGGGCUUUUGGACCC CGUGGCCCGGGUUCGAAUCCCGGCCGGGCUACCA
Pyrococcus abyssi (CUG)	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUCUGGAUCC CGCGACCGGGGUUCGAAUCCCCGCGGGGCUACCA
Pyrococcus abyssi (UUG)	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUUUUGGAUCC CGCGACCCGGGUUCGAAUCCCGGCGGGGCUACCA
Pyrococcus furiosus (CUG)	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUCUGGAUCC CGCGACCGGGGUUCGAAUCCCCGCGGGGCUACCA
Pyrococcus furiosus (UUG)	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUUUUGGAUCC CGCGACCCGGGUUCGAAUCCCGGCGGGGCUACCA
Pyrococcus horikoshii (CUG)	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUCUGGAUCC CGCGACCGGGGUUCGAAUCCCCGCGGGGCUACCA
Pyrococcus horikoshii (UUG)	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUUUUGGAUCC CGCGACCCGGGUUCGAAUCCCGGCGGGGCUACCA
Sulfolobus solfataricus (CUG)	AGCCGGGUAGUCUAGUGGUCAAGGAUCCAGGGCUCUGGCCCC UGGGACCAGGGUUCGAAUCCCUGCCCGGCUACCA
Sulfolobus solfataricus (UUG)	AGCCGGGUAGUCUAGUGGUCAAGGAUCCAGGGCUUUUGGCCCC UGGGACCAGGGUUCGAAUCCCUGCCCGGCUACCA
Sulfolobus tokodaii (CUG)	AGCCGGGUCGUCUAGUGGUCAAGGAUCGAGGGCUCUGGCCCC UCGGACCUGGGUUCAAAUCCCAGCCCGGCUACCA
Sulfolobus tokodaii (UUG)	AGCCGGGUCGUCUAGUGGUCAAGGAUCGAGGGCUUUUGGCCCC UCGGACCUGGGUUCAAAUCCCAGCCCGGCUACCA
Thermoplasma acidophilum (CUG)	AGCCCUGUGGUGUAGUGGCCAAGCAUUAUGGGCUCUGGACCC AUCGACGGCAGUUCGAAUCUGCCCAGGGCUACCA
Thermoplasma acidophilum (UUG)	AGCCCUGUGGUGUAGUGGACAAGCAUUUUGGACUUUUGGAUUC CAAAGACGGCAGUUCGAAUCUGCCCAGGGCUACCA

-69-

tRNA (anticodon)	Sequence[1]
Thermoplasma volcanium (CUG)	AGCCCUGUGGUGUAGCGGCCAAGCAUUAUGGGCUCUGGACC AUCGACGGCAGUUCGAAUCUGCCCAGGGCUACCA
Thermoplasma volcanium (UUG)	AGCCCUGUGGUGUAGUGGACAAGCAUUUUGGACUUUGGAUCC AACGACGGCAGUUCGAAUCUGCCCAGGGCUACCA
Consensus (YUG)[2]	AGCCCCGUGGUGUAGYGGCCAAGCAUGCGGGRCUYUGGAYCC CGCGACCGSGGUUCGAAUCCSCGCGGGGCUACCA
AQ(GU) (CUA)[3]	AGCCCCGUGGUGUAGcGGCCAAGCAUGCGGGgCUcUaaAcCCCGC GACCGgGGUUCGAAUCCcCGCGGGGCUACCA
AQ(GC) (CUA)[3]	AGCCCCGUGGUGUAGcGGCCAAGCAcGCGGGgCUcUaaAcCCCGC GACCGgGGUUCGAAUCCcCGCGGGGCUACCA

[1] Bases indicated by underlining deviate from the identity of their consensus bases. Bases highlighted in black correspond to positions at which two different bases predominate with approximately equal frequency. Anticodon positions are boxed.

[2] Bases that are in boldface correspond to positions involved in non-canonical base pairs (Figure 1, Panel A). Y corresponds to bases C or U; R corresponds to bases A or G; S corresponds to bases C or G.

[3] Bases in lowercase correspond to positions that were varied from the consensus sequence in the design of a prospective orthogonal tRNA (Figure 1, Panel B).

[0182]   *Orthogonality of tRNAs:* To test the activity and cross-reactivity of the AE(GU), AQ(GU), and AQ(GC) tRNAs, their genes were inserted into plasmid pACKO-A184TAG. *E. coli* cells were transformed with the resulting plasmids, pAC-AE(GU), pAC-AQ(GU), and pAC-AQ(GC), and the ampicillin IC_{50} values for the cells were assayed. The tRNA-expressing pAC-based plasmids carry a β-lactamase reporter gene containing an amber mutation at the permissive site Ala184, which serves as a useful reporter for amber codon suppression. Acylation of an amber suppressor tRNA causes the incorporation of the activated amino acid at position 184 of β-lactamase, resulting in suppression of the amber stop codon, production of active full-length β-lactamase and resistance to ampicillin. Because an orthogonal tRNA should not cross-react with endogenous aminoacyl-tRNA synthetases, strains transformed with an orthogonal suppressor tRNA alone should survive only on low concentrations of ampicillin. To assess the orthogonality of designed tRNAs, *E. coli* transformants were plated on LB agar plates containing varying concentrations of ampicillin (**Figure 2**). Cells expressing the AQ(GU) and AQ(GC) tRNAs exhibited ampicillin IC_{50} values of 60 ± 12 and 20 ± 4 μg/mL, respectively. Cells expressing the AE(GU) tRNA exhibited a relatively lower ampicillin IC_{50} value of 5 ± 1 μg/mL, suggesting that this tRNA is the least cross-reactive of the three.

[0183]   *Cloning and assaying archaeal glutamyl-tRNA synthetases in E. coli:* Based on previous experiments (*see, e.g.*, Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal*

suppression, Biochemistry, Aug 19;42(32):9598-9608), it was anticipated that the AE(GU),

AQ(GU) and AQ(GC) tRNAs might serve as substrates for archaeal glutamyl-tRNA

synthetases. Suppression efficiencies corresponding to various synthetases-tRNA pair

combinations were assessed using the ampicillin IC_{50} assay. If the designed tRNAs are

5   recognized by the archaeal synthetases, then *E. coli* cells co-expressing the synthetase-

tRNA pairs should exhibit ampicillin IC_{50} values greater than those displayed by cells

expressing tRNAs alone. The AfERS, ApERS, MmERS, MtERS, PhERS, and SsERS genes

were each inserted into the vector pKQ, which can be co-maintained in *E. coli* in

combination with the tRNA-expressing plasmids. Following co-transformation of *E. coli*

10   with all pair-wise combinations of synthetase- and tRNA-expressing plasmids, cells were

grown on LB agar plates containing varying concentrations of ampicillin. Cells expressing

synthetase-AQ(GU) or synthetase-AQ(GC) combinations exhibited IC_{50} values equivalent

to those of cells expressing AQ(GU) alone or AQ(GC) alone, respectively, suggesting that

neither of these tRNAs is recognized by any of the archaeal glutamyl-tRNA synthetases

15   when co-expressed in *E. coli*. In contrast, cells expressing combinations of synthetase and

AE(GU) tRNA exhibited ampicillin IC_{50} values 2–20-fold higher than those displayed by

cells expressing AE(GU) alone (**Figure 3**). The greatest IC_{50} value differences were

exhibited by cells co-expressing the AE(GU) tRNA together with MtERS (6 fold), AfERS

(7 fold), PhERS (12 fold), and MmERS (20 fold) . Cells co-expressing the best pair,

20   MmERS and AE(GU), exhibited an IC_{50} value of 100 µg/mL, a value still 4.5-fold lower

than that displayed by the *M. jannaschii*-derived tyrosine pair (*see, e.g.*, Wang,L., Brock,A.,

Herberich,B. and Schultz,P.G. (2001) *Expanding the genetic code of Escherichia coli.*

Science, 292:498-500) and 10-fold lower than that displayed by the *M.*

thermoautotrophicum-derived leucine pair (*see, e.g.*, Anderson JC, & Schultz PG, (2003),

25   *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base,*

amber, and opal suppression, Biochemistry, Aug 19;42(32):9598-9608). These results

indicate that the AE(GU) tRNA combined with MtERS, AfERS, PhERS or MmERS

constitute active but inefficient orthogonal pairs in *E. coli*.

[0184]   *Selection and characterization of an optimized consensus tRNA^{Glu}.* In order

30   to increase the efficiency of the archaeal glutamic acid orthogonal pairs, improved variants

of the AE(GU) tRNA were sought. The presence of the non-canonical G10-U28 base pair

was regarded as a possible explanation for the low amber-suppression efficiency exhibited

by synthetase-AE(GU) pairs. An attempt to investigate the importance of this G-U base pair had already been made with the construction of the two tRNA^{Gln}-based consensus tRNAs, AQ(GU) and AQ(GC), which differed only with respect to this base pair. However, because both AQ(GU) and AQ(GC) lacked the ability to be aminoacylated by archaeal glutamyl-

5   tRNA synthetases in *E. coli*, the importance of the G-U base pair was still unknown. To further address this issue, a small library of variants of the consensus tRNA^{Glu}, AE(GU), was constructed in which nucleotides in the 10 and 28 positions were randomized (**Figure 4, Panel A**). This "AE(NN)" tRNA library was constructed by PCR using a degenerate oligonucleotide primer (*see* Materials and Methods), and inserted into the tRNA cloning site

10   of plasmid pACKO-A184TAG, to afford the pAC-AE(NN) plasmid library. *E. coli* cells co-transformed with plasmid pKQ-MmERS and plasmid library pAC-AE(NN) were then plated on a series of LB agar plates containing varying concentrations of ampicillin. Bacterial colonies were observed on plates containing ampicillin concentrations as high as 500 μg/mL, indicating that at least one variant of the AE(GU) tRNA exhibited enhanced

15   amber-suppression efficiency compared to the parent when co-expressed with MmERS in *E. coli*. Ten bacterial colonies were picked from the 500-μg/mL ampicillin plate, and the pAC-AE(NN) plasmids were sequenced. All ten of the clones contained a G10-C28 base pair (**Figure 4, Panel B**), indicating that this "AE(GC)" tRNA is the most active variant within the AE(NN) library.

20   **[0185]**   To measure the amber suppression efficiency of the selected AE(GC) tRNA, *E. coli* cells carrying either pAC-AE(GC) alone or in combination with plasmid pKQ-AfERS, pKQ-MmERS, pKQ-MtERS, or pKQ-PhERS were plated on a series of LB agar plates containing varying concentrations of ampicillin. In all cases, *E. coli* cells co-expressing the AE(GC) tRNA were found to exhibit ampicillin IC$_{50}$ values 4–5-fold greater

25   than those of cells co-expressing the AE(GU) tRNA (**Figure 4, Panel C**). Cells co-expressing the MmERS and AE(GC) pair exhibited an ampicillin IC$_{50}$ value of 450 μg/mL, 4.5-fold greater than that displayed by cells co-expressing the MmERS and AE(GU) combination. Similarly, cells coexpressing the PhERS and AE(GC) pair exhibited an IC$_{50}$ value of 212 μg/mL, ~3.5-fold greater than those co-expressing PhERS and AE(GU). The

30   ampicillin IC$_{50}$ value for cells expressing the AE(GC) tRNA alone also increased 4-fold relative to cells expressing the AE(GU) tRNA alone (20 μg/mL versus 5 μg/mL ampicillin), suggesting that the U28 to G replacement enhanced the overall amber suppression

efficiency of the tRNA. Without being limited to one theory, such an enhancement may

have resulted from improved *E. coli* processing of the AE(GC) tRNA compared to AE(GU)

(*see, e.g.*, Anderson JC, & Schultz PG, (2003), *Adaptation of an orthogonal archaeal*

leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression, Biochemistry,

5   Aug 19;42(32):9598-9608). Alternatively, the increased IC_{50} value could be due to an

increased rate of background aminoacylation of the AE(GC) tRNA compared to AE(GU).

[0186]   *Orthogonality of archaeal glutamyl-tRNA synthetases in E. coli*: The

ampicillin IC_{50} measurements described above provide insights into the orthogonality of

consensus-designed tRNAs, but provide no information about the orthogonality of archaeal

10   glutamyl-tRNA synthetases in *E. coli*. To investigate synthetase orthogonality, the AfERS,

MmERS, MtERS, and PhERS genes were inserted into a protein overexpression vector. *E.*

coli glutamyl-tRNA synthetase (EcERS) was inserted into the same vector for use as a

control. Synthetases were overexpressed in *E. coli*, purified to homogeneity, and used in an

in vitro aminoacylation assay with whole tRNA prepared from either *E. coli* or the

15   halophilic archaebacterium *Halobacterium sp. NRC-1*. As expected, EcERS robustly

aminoacylated whole *E. coli* tRNA (14,000 pmol/L) and exhibited virtually no activity

toward whole archaeal tRNA (1.7 pmol/L; **Figure 5**). In contrast, none of the four archaeal

synthetases exhibited measurable aminoacylation activity with whole *E. coli* tRNA (≤ 2.8

pmol/L), while two of the four synthetases actively aminoacylated whole archaeal tRNA

20   (530 and 1100 pmol/L for MtERS and PhERS, respectively). In these experiments, it is

unclear why AfERS and MmERS could not aminoacylate whole halobacterial tRNA *in vitro*

(≤ 5 pmol/L). Without being limited to one theory, it is possible that the enzyme purification

and/or assay conditions employed were deleterious to the aminoacylation activity of AfERS

and MmERS. Another possibility is that subtle differences in tRNA identity elements might

25   preclude recognition of a given archaeal tRNA^{Glu} by all archaeal glutamyl-tRNA

synthetases. To test the latter, the activity of MmERS was investigated using the *in vitro*

aminoacylation assay with whole tRNA derived from *E. coli* expressing the AE(GC) tRNA.

Again, no activity was observed, suggesting that the C-terminally His-tagged AfERS and

MmERS enzymes are not active following purification. The above data demonstrate the

30   orthogonality of MtERS and PhERS, and the likely orthogonality of AfERS and MmERS.

[0187]   The consensus sequence strategy for the design of orthogonal tRNAs was

successfully applied to the development of an orthogonal tRNA based on the family of

tRNA^{Glu} sequences from archaea. The initial candidate tRNA^{Glu}-based tRNA, AE(GU), was found to be highly orthogonal in *E. coli*, with cells expressing the tRNA alone displaying an ampicillin IC$_{50}$ value of just 5 µg/mL. However, the AE(GU) tRNA inefficiently suppressed amber codons when expressed in combination with archaeal glutamyl-tRNA synthetases in

5   *E. coli*. Cells containing this synthetase-tRNA pair displayed an ampicillin IC$_{50}$ value of only 60 and 100 µg/mL for the two best synthetases, PhERS and MmERS, respectively. In order to improve the amber suppression efficiency of the AE(GU) tRNA, a small library of variants, randomized at the 10 and 28 positions, was constructed. Library members were selected for the ability to efficiently suppress an amber stop codon in *E. coli* when

10 expressed in combination with MmERS. The resulting AE(GC) tRNA, when paired with PhERS and MmERS, suppress amber codons with high efficiency. Cells co-expressing this tRNA with PhERS and MmERS display ampicillin IC$_{50}$ values of 212 and 450 µg/mL, respectively.

[0188]   An orthogonal tRNA based on the tRNA^{Gln} family of sequences from the
15 same organisms was not developed with the consensus sequence strategy. Without being limited to one theory, members of the family of archaeal tRNA^{Gln} sequences, in general, deviate more from their consensus tRNA than do tRNA^{Glu} sequences from theirs (Tables 1 and 2). This deviation suggests that individuals within the family of tRNA^{Gln} may cluster less tightly in their evolutionary relatedness compared to tRNAs within the archaeal

20 tRNA^{Glu} family. Moreover, the degree of deviation within the tRNA^{Gln} family may prevent individual synthetases from productively recognizing the consensus tRNA. A second possible explanation for the failure of the tRNA^{Gln}-based consensus design is that the AQ(GU) and AQ(GC) tRNAs may lack unknown elements that permit efficient processing in *E. coli*. *E. coli* cells expressing the AQ(GU) and AQ(GC) tRNAs without co-expressed

25 archaeal glutamyl-tRNA synthetases exhibit relatively relatively high ampicillin IC$_{50}$ values compared to cells expressing the AE(GU) tRNA or no exogenous tRNA (**Figure 2**). Without being limited to one theory, these data suggest that the AQ(GU) and AQ(GC) tRNAs are processed to at least some extent in *E. coli* and that their inability to be charged by archaeal glutamyl-tRNA synthetases in *E. coli* is due to the fact that they are not

30 recognized as substrates.

[0189]   The glutamyl-tRNA synthetase from *P. horikoshii* functions as part of an orthogonal tRNA synthetase and amber suppressor tRNA pair in *E. coli*. The development

of this pair is a significant step in the process of incorporating unnatural amino acids into

proteins *in vivo*. Synthetase variants can be evolved to selectively recognize unnatural

amino acids and catalyze their condensation with the orthogonal AE(GC) tRNA. The

evolved synthetases can then be used in unnatural amino acid mutagenesis of target proteins

5   in *E. coli*.

[0190]   Without being limited to one theory, it is likely that the amber suppression

efficiency of the orthogonal synthetase and tRNA pair in *E. coli* is an important determinant

of synthetase evolvability. Amber suppression efficiency, in turn, is determined by several

kinetic parameters, including the rate of synthetase and tRNA production and processing in

10   *E. coli* and the catalytic efficiency of tRNA aminoacylation. Altering the amino acid

substrate specificity of a synthetase through the variation of active site residues likely

reduces amber suppression efficiency by affecting either the k_{cat} or K_M of the variant

enzyme (*see, e.g.,* Wang,L., Brock,A., Herberich,B. and Schultz,P.G. (2001) *Expanding the

genetic code of Escherichia coli.* Science, 292:498-500; and, Agou,F., Quevillon,S.,

15   Kerjan,P. and Mirande,M. (1998) *Switching the amino acid specificity of an aminoacyl-

tRNA synthetase.* Biochemistry, 11:11309-11314). Thus, the suppression efficiency of the

parent synthetase-tRNA pair should be high enough to withstand a reduction in catalytic

efficiency that can accompany protein evolution and allow the selection criteria to be met.

[0191]   Amber suppression efficiencies can be determined by measurement of

20   ampicillin IC_{50} values for *E. coli* bearing a synthetase-tRNA pair. The only orthogonal

synthetase that has been evolved to accept unnatural amino acids in *E. coli*, the tyrosyl-

tRNA synthetase from *M. jannaschii*, exhibits an ampicillin IC_{50} value of 440 μg/mL (*see,

e.g.,* Wang,L., Brock,A., Herberich,B. and Schultz,P.G. (2001) *Expanding the genetic code

of Escherichia coli.* Science, 292:498-500). In contrast, the glutamine pair derived from *S.

25   cerevisiae*, for which variants capable of accepting unnatural amino acids have not yet been

found, displays an ampicillin IC_{50} value of only 140 μg/mL (*see, e.g.,* Pastrnak,M.,

Magliery,T.J. and Schultz,P.G. (2000) *A new orthogonal suppressor tRNA/aminoacyl-tRNA

synthetase pair for evolving an organism with an expanded genetic code.* Helvetica Chimica

Acta, 83:2277-2286). The difference in suppression efficiency between these two pairs may

30   explain the apparent difference in the evolvability of their synthetases. On the basis of these

data, PhERS, which, together with the AE(GC) tRNA, exhibits an amber suppression

efficiency in *E. coli* ~2-fold less than that of the tyrosyl-tRNA synthetase, can be amenable

to selection experiments to alter amino acid specificity. MmERS, which has an ampicillin IC$_{50}$ value roughly comparable to that of the tyrosyl-tRNA synthetase, can also be amenable to selection. Besides catalytic efficiency, however, other factors are likely to play a role in synthetase evolvability. The extensive hydrogen bonding network within the active site of

5   the glutamine synthetase, for example, may affect the enzyme's evolvability. Our understanding of the relative importance of activity versus other, perhaps more subtle, features will improve our ability to generate and evolve additional orthogonal pairs.

[0192]   The *M. jannaschii*-derived tyrosine orthogonal pair has been evolved to recognize a variety of unnatural amino acids containing aryl side chains. Efforts to develop

10   variants of this synthetase that accept more diverse chemical structures are in progress. The availability of additional orthogonal pairs opens the possibility of expanding the diversity of unnatural amino acids that can be incorporated in, e.g., *E. coli*. The glutamic acid orthogonal pair described here, for example, can facilitate the development of synthetase variants that recognize negatively charged amino acids. This involves the generation of a

15   large library of synthetase variants containing random amino acids substitutions at specific residues involved in binding the natural amino acid. Therefore, the availability of high-resolution structural information about the amino acid substrate-binding site is helpful for the directed evolution of a synthetase. Although the x-ray crystal structure of the glutamyl-tRNA synthetase from *M. mazei* has not yet been solved, the structure of the homologous

20   enzyme from *Thermus thermophilus* is available (*see, e.g.,* Nureki,O., Vassylyev,D.G., Katayanagi et al., (1995) *Architectures of class-defining and specific domains of glutamyl-tRNA synthetase.* Science, 31:1958-1965; and, Sekine et al., (2001) *Structural basis for. anticodon recognition by discriminating glutamyl-tRNA synthetase.* Nat. Struct. Biol., 8:203-206). This information is supplemented by co-crystal structural data of the

25   glutaminyl-tRNA synthetase from *E. coli* with a glutaminyl-adenylate analog (*see, e.g.,* Rath et al., (1998) *How glutaminyl-tRNA synthetase selects glutamine.* Structure, 6:439-449). Although they recognize somewhat different amino acid structures, the amino acid binding pockets of these enzymes are superimposable.

[0193]   For many applications, the ability to incorporate multiple distinct unnatural

30   amino acids into a protein *in vivo* is highly desirable. For example, the incorporation of two different fluorescent amino acids into a protein molecule would enable the use of fluorescence resonance energy transfer (FRET) for studying the conformational dynamics

of proteins *in vivo*. The incorporation of two distinct unnatural amino acids into a protein requires the use of two mutually orthogonal synthetases that have been evolved to accept different unnatural amino acids. The development of an orthogonal glutamyl-tRNA synthetase and tRNA pair further advances the potential that multiple distinct unnatural

5   amino acids can be incorporated into a single protein chain in living cells. Moreover, the archaeal species represent a rich source of orthogonal pairs for use in unnatural amino acid mutagenesis in, e.g., *E. coli*. These results provide insight into the rational design of efficiently-charged orthogonal tRNAs.

EXAMPLE 2: EXEMPLARY GLUTAMYL O-RSs AND GLUTAMYL O-tRNAs.

10   **[0194]**   An exemplary O-tRNA comprises, e.g., SEQ ID NO.: 67 (and others herein) (*See* Table 3). Exemplary O-RSs include, e.g., SEQ ID NOs.: 69, 73, 75 and 77 (and others herein) (*See* Table 3). Exemplary polynucleotides that encode O-RSs or portions thereof include, e.g., SEQ ID NOs.: 68, 72, 74 and 76 (and others herein).

[0195]   Further details of the invention, and in particular experimental details, can be

15   found in Anderson, John Christopher, "Pathway Engineering of the Expanding Genetic Code," Ph.D. Dissertation, The Scripps Research Institute [2003].

[0196]   It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of

20   this application and scope of the appended claims.

[0197]   While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above

25   can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

30

TABLE 3: SEQUENCES

SEQ ID:	Label	SEQUENCE
SEQ ID: 1	*Aeropyrum pernix* (CUC) Glutamic acid tRNA	GCCGCGGUAGUAUAGCCCGGCCCAGUAUGCGGGCCUCUCGAGCCC GUGACCCGGGUUCAAAUCCCGGCCGCGGCACCA
SEQ ID: 2	*Aeropyrum pernix* (UUC) Glutamic acid tRNA	GCCGCGGUAGUAUAGCCCGGCCCAGUAUGCGGGCCUUUCGAGCCC GUGACCCGGGUUCAAAUCCCGGCCGCGGCACCA
SEQ ID: 3	*Archaeogl obus fulgidus* (CUC) Glutamic acid tRNA	GCUCCGGUGGUGUAGCCCGGCCAAUCAUACGGGACUCUCGAUCCC GUGACCCGGGUUCAAAUCCCGGCCGGAGCACCA
SEQ ID: 4	*Archaeogl obus fulgidus* (UUC) Glutamic acid tRNA	GCUCCGGUGGUGUAGCCCGGCCAAUCAUUCCGGCCUUUCGAGCCG GCGACCCGGGUUCAAAUCCCGGCCGGAGCACCA
SEQ ID: 5	*Halobacte rium sp.* (CUC) Glutamic acid tRNA	GCUCCGUUGGUGUAGUCCGGCCAAUCAUCUUGCCCUCUCACGGCA AGGACUAGGGUUCAAAUCCCUGACGGAGCACCA
SEQ ID: 6	*Halobacte rium sp.* (UUC) Glutamic acid tRNA	GCUCGGUUGGUGUAGUCCGGCCAAUCAUGUUGGCCUUUCGAGCCG ACGACCAGGGUUCAAAUCCCUGACCGAGCACCA
SEQ ID: 7	*Methanoba cterium thermauto trophicus* (UUC) Glutamic acid tRNA	GCUCCGGUAGUGUAGUCCGGCCAAUCAUUUCGGCCUUUCGAGCCG AAGACUCGGGUUCAAAUCCCGGCCGGAGCACCA
SEQ ID: 8	*Methanoco ccus jannaschi i* (UUC) Glutamic acid tRNA	GCUCCGGUGGUGUAGUCCGGCCAAUCAUGCGGGCCUUUCGAGCCC GCGACCCGGGUUCAAAUCCCGGCCGGAGCACCA
SEQ ID: 9	*Methanoco ccus maripalud is* (UUC) Glutamic acid tRNA	GCUCCAGUGGUGUAGUCCGGCCAAUCAUCCGGCCCUUUCGAGGCC GGGACUCGGGUUCAAAUCCCGGCUGGAGCACCA
SEQ ID: 10	*Methanosa rcina mazei* (CUC)	GCUCCGAUAGUGUAGCUCGGCCAAUCAUUCAGGACUCUCACUCCU GCGACUGGGGUUCAAAUCCCCAUCGGAGCACCA

	Glutamic acid tRNA	
SEQ ID: 11	*Methanosarcina mazei* (UUC) Glutamic acid tRNA	GCUCCGGUAGUGUAGUCCGGCCAAUCAUUCCGGCCUUUCGAGCCG AAGACUCGGGUUCGAAUCCCGGCCGGAGCACCA
SEQ ID: 12	*Pyrobaculum aerophilum* (CUC) Glutamic acid tRNA	CAGCCGGUAGUCUAGCCCGGACAAGGAUGCGGGCCUCUCGAGCCC GUGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 13	*Pyrococcus abyssi* (CUC) Glutamic acid tRNA	GCCCCGGUGGUGUAGCCCGGUCAAUCAUGCGGGACUCUCGAUCCC GCGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 14	*Pyrococcus abyssi* (UUC) Glutamic acid tRNA	GCCCCGGUGGUGUAGCCCGGUCAAACAUGCGGGCCUUUCGAGCCC GCGCCCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 15	*Pyrococcus furiosus* (CUC) Glutamic acid tRNA	GCCCCGGUGGUGUAGCCCGGUCAAUCAUGCGGGACUCUCGAUCCC GCGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 16	*Pyrococcus furiosus* (UUC) Glutamic acid tRNA	GCCCCGGUGGUGUAGCCCGGUCAAACAUGCGGGCCUUUCGAGCCC GCGCCCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 17	*Pyrococcus horikoshii* (CUC) Glutamic acid tRNA	GCCCCGGUGGUGUAGCCCGGUCAAUCAUGCGGGACUCUCGAUCCC GCGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 18	*Pyrococcus horikoshii* (UUC) Glutamic acid tRNA	GCCCCGGUGGUGUAGCCCGGCCAAACAUGCGGGCCUUUCGAGCCC GCGCCCCGGGUUCAAAUCCCGGCCGGGGCACCA.
SEQ ID: 19	*Sulfolobus solfataricus* (CUC) Glutamic acid tRNA	GCCGCGGUGGUUUAGCCCGGUC- AGAAUGCGGGCCUCUCGAGCCCGUGACCCGGGUUCAAAUCCCGGC CGCGGCACCA
SEQ ID: 20	*Sulfolobus solfataricus* (UUC) Glutamic	GCCGCGGUGGUUUAGCCCGGUC- AGAAUGCGGGCCUUUCGAGCCCGUGACCCGGGUUCAAAUCCCGGC CGCGGCACCA

	acid tRNA	
SEQ ID: 21	*Sulfolobus tokodaii* (CUC) Glutamic acid tRNA	GCCGCGGUAGUAUAGCCCGGUC-AGUAUGCGGGCCUCUCAAGCCCGUGGCCCGGGUUCAAAUCCCGGCCGCGGCACCA
SEQ ID: 22	*Sulfolobus tokodaii* (UUC) Glutamic acid tRNA	GCCGCGGUAGUAUAGCCCGGUCAAGCACGCGGGCCUUUCGAGCCCGUGACCCGGGUUCAAAUCCCGGCCGCGGCACCA
SEQ ID: 23	*Thermoplasma acidophilum* (CUC) Glutamic acid tRNA	GCUCCGGUGGUGUAGUCCGGCCAAGCAUUAGGGACUCUCAAUCCCCCGACUCGGGUCCAAAUCCCGACCGGAGCACCA
SEQ ID: 24	*Thermoplasma acidophilum* (UUC) Glutamic acid tRNA	GCUCCGGUGGUGUAGUCCGGACAAGCAUAUCGGCCUUUCGAGCCGACGACCUGGGUCCAAAUCCCGGCCGGAGCACCA
SEQ ID: 25	*Thermoplasma volcanium* (CUC) Glutamic acid tRNA	GCUCCGGUGGUGUAGUCCGGCCAAGCAUUAGGGACUCUCAAUCCCCCGACUCGGGUCCAAAUCCCGACCGGAGCACCA
SEQ ID: 26	*Thermoplasma volcanium* (UUC) Glutamic acid tRNA	GCUCCGGUGGUGUAGUCCGGACAAGCAUAUCGGCCUUUCGAGCCGACGACCUGGGUCCAAAUCCCGGCCGGAGCACCA
SEQ ID: 27	Consensus (YUC) Glutamic acid tRNA	GCYCCGGUGGUGUAGCCCGGCCAAKCAUGCGGGCCUYUCGAGCCCGCGACCCGGGUUCAAAUCCCGGCCGGRGCACCA
SEQ ID: 28	AE (GU) (CUA) Glutamic acid tRNA	GCCCCGGUGGUGUAGCCCGGCCAAGCAUGCGGGCCUCUAAAGCCCGCGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 29	*Aeropyrum pernix* (CUG) Glutamine tRNA	AGCCGGGUCGUCUAGCGGCCCAGGAUGCGGGGCUCUGGCCCCCGUGACCGGGGUUCGAAUCCCCGCCCGGCUACCA
SEQ ID: 30	*Aeropyrum pernix* (UUG) Glutamine tRNA	AGCCGGGUCGUCUAGCGGCCCAGGAUGCGGGGUUUUGGCCCCCGUGACCCGGGUUCGAAUCCCGGCCCGGCUACCA
SEQ ID: 31	*Archaeoglobus fulgidus* (CUG)	AGCCCCGUGGGGUAGCGGUCAAUCCUGCCGGACUCUGGAUCCGGCGACGCCGGUUCGAAUCCGGCCGGGGCUACCA

	Glutamine tRNA	
SEQ ID: 32	*Archaeogl obus fulgidus* (UUG) Glutamine tRNA	AGCCCCGUGGUGUAGCGGUCAAUCAUGCGGGCCUUUGGAGCCCGC GACCGGGGUUCGAAUCCCCGCGGGGCUACCA
SEQ ID: 33	*Halobacte rium sp.* (CUG) Glutamine tRNA	AGUCCCGUAGGGUAGUGGCCAAUCCUGAAGCCUUCUGGGGGCUUC GACGGAAGUUCGAAUCUUCCCGGGACUACCA
SEQ ID: 34	*Halobacter sp.* (UUG) Glutamine tRNA	AGUCCCGUGGUGUAGUGGCCAAUCAUAUGGGCCUUUGGAGCCCAC GACGGCGGUUCGAAUCCGCCCGGGACUACCA
SEQ ID: 35	*Methanoba cterium thermauto trophicus* (CUG) Glutamine tRNA	AGUCCCGUGGGGUAAUGGCA- AUCCUGAUGGACUCUGGAUCCAUCGAUAGCGGUUCGACUCCGCUC GGGACUACCA
SEQ ID: 36	*Methanoba cterium thermauto trophicus* (UUG) Glutamine tRNA	AGUCCCGUGGGGUAGUGGUA- AUCCUGCUGGGCUUUGGACCCGGCGACAGCGGUUCGACUCCGCUC GGGACUACCA
SEQ ID: 37	*Methanoco ccus jannaschi i* (UUG) Glutamine tRNA	AGCCCGGUGGUGUAGUGGCCUAUCAUCCGGGGCUUUGGACCCCGG GACCGCGGUUCGAAUCCGCGCCGGGCUACCA
SEQ ID: 38	*Methanoco ccus maripalud is* (UUG) Glutamine tRNA	AGCCCAGUAGUGUAGUGGCCAAUCAUCCGUGCCUUUGGAGCAUGG GACCGCGGUUCGAAUCCGCGCUGGGCUACCA
SEQ ID: 39	*Methanopy rus kandleri* (UUG) Glutamine tRNA	AGGGCCGUGGGGUAGCGGUCUAUCCUGCGGGGCUUUGGACCCCGC GACCCCGGUUCAAAUCCGGGCGGCCCUACCA
SEQ ID: 40	*Methanosa rcina mazei* (CUG) Glutamine tRNA	AGCCCGGUAGUGUAGUGGUCAAUCAUGCGGGACUCUGGAUCCUGC AACCUCGGUUCGAAUCCGUGCCGGGCUACCA
SEQ ID: 41	*Methanosa rcina mazei* (UUG)	AGUCCUGUAGGGUAGUGGUCAAUCCUUCGGGCCUUUGGAGCCCGG GACAGCGGUUCGAAUCCGCUCAGGACUACCA

	Glutamine tRNA	
SEQ ID: 42	*Pyrobaculum aerophilum* (CUG) Glutamine tRNA	AGCCCGGUCGUCUAGCGGCC-AGGAUGCGGGGCUCUGGACCCCGUGGCCCGGGUUCGAAUCCCGGCCGGGCUACCA
SEQ ID: 43	*Pyrobaculum aerophilum* (UUG) Glutamine tRNA	AGCCCGGUCGUCUAGCGGCCAAGGAUGCGGGGCUUUGGACCCCGUGGCCCGGGUUCGAAUCCCGGCCGGGCUACCA
SEQ ID: 44	*Pyrococcus abyssi* (CUG) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUCUGGAUCCCGCGACCGGGGUUCGAAUCCCCGCGGGGCUACCA
SEQ ID: 45	*Pyrococcus abyssi* (UUG) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUUUGGAUCCCGCGACCCGGGUUCGAAUCCCGGCGGGGCUACCA
SEQ ID: 46	*Pyrococcus furiosus* (CUG) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUCUGGAUCCCGCGACCGGGGUUCGAAUCCCCGCGGGGCUACCA
SEQ ID: 47	*Pyrococcus furiosus* (UUG) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUUUGGAUCCCGCGACCCGGGUUCGAAUCCCGGCGGGGCUACCA
SEQ ID: 48	*Pyrococcus horikoshii* (CUG) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUCUGGAUCCCGCGACCGGGGUUCGAAUCCCCGCGGGGCUACCA
SEQ ID: 49	*Pyrococcus horikoshii* (UUG) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGACUUUGGAUCCCGCGACCCGGGUUCGAAUCCCGGCGGGGCUACCA
SEQ ID: 50	*Sulfolobus solfataricus* (CUG) Glutamine tRNA	AGCCGGGUAGUCUAGUGGUCAAGGAUCCAGGGCUCUGGCCCCUGGGACCAGGGUUCGAAUCCCUGCCCGGCUACCA
SEQ ID: 51	*Sulfolobus solfataricus* (UUG) Glutamine	AGCCGGGUAGUCUAGUGGUCAAGGAUCCAGGGCUUUGGCCCCUGGGACCAGGGUUCGAAUCCCUGCCCGGCUACCA

	tRNA	
SEQ ID: 52	*Sulfolobus tokodaii* (CUG) Glutamine tRNA	AGCCGGGUCGUCUAGUGGUCAAGGAUCGAGGGCUCUGGCCCCUCG GACCUGGGUUCAAAUCCCAGCCCGGCUACCA
SEQ ID: 53	*Sulfolobus tokodaii* (UUG) Glutamine tRNA	AGCCGGGUCGUCUAGUGGUCAAGGAUCGAGGGCUUUGGCCCCUCG GACCUGGGUUCAAAUCCCAGCCCGGCUACCA
SEQ ID: 54	*Thermoplasma acidophilum* (CUG) Glutamine tRNA	AGCCCUGUGGUGUAGUGGCCAAGCAUUAUGGGCUCUGGACCCAUC GACGGCAGUUCGAAUCUGCCCAGGGCUACCA
SEQ ID: 55	*Thermoplasma acidophilum* (UUG) Glutamine tRNA	AGCCCUGUGGUGUAGUGGACAAGCAUUUUGGACUUUGGAUCCAAA GACGGCAGUUCGAAUCUGCCCAGGGCUACCA
SEQ ID: 56	*Thermoplasma volcanium* (CUG)	AGCCCUGUGGUGUAGCGGCCAAGCAUUAUGGGCUCUGGACCCAUC GACGGCAGUUCGAAUCUGCCCAGGGCUACCA
SEQ ID: 57	*Thermoplasma volcanium* (UUG) Glutamine tRNA	AGCCCUGUGGUGUAGUGGACAAGCAUUUUGGACUUUGGAUCCAAC GACGGCAGUUCGAAUCUGCCCAGGGCUACCA
SEQ ID: 58	Consensus (YUG) Glutamine tRNA	AGCCCCGUGGUGUAGYGGCCAAGCAUGCGGGRCUYUGGAYCCCGC GACCGSGGUUCGAAUCCSCGCGGGGCUACCA
SEQ ID: 59	AQ(GU) (CUA) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGGCUCUAAACCCCGC GACCGGGGUUCGAAUCCCCGCGGGGCUACCA
SEQ ID: 60	AQ(GC) (CUA) Glutamine tRNA	AGCCCCGUGGUGUAGCGGCCAAGCACGCGGGGCUCUAAACCCCGC GACCGGGGUUCGAAUCCCCGCGGGGCUACCA
SEQ ID: 61	Glu consensus tRNA	GCYCCGGUGGUGUAGCCCGGCCAAKCAUGCGGGCCUYUCGAGCCC GCGACCCGGGUUCAAAUCCCGGCCGGRGCACCA
SEQ ID: 62	Gln consensu tRNA	AGCCCCGUGGUGUAGYGGCCAAGCAUGCGGGRCUYUGGAYCCCGC GACCGSGGUUCGAAUCCSCGCGGGGCUACCA
SEQ ID: 63	AE(GU) tRNA	GCCCCGGUGGUGUAGCCCGGCCAAGCAUGCGGGCCUCUAAAGCCC GCGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 64	AQ(GU) tRNA	AGCCCCGUGGUGUAGCGGCCAAGCAUGCGGGGCUCUAAACCCCGC GACCGGGGUUCGAAUCCCCGCGGGGCUACCA

SEQ ID: 65	AQ(GC) tRNA	AGCCCCGUGGUGUAGCGGCCAAGCACGCGGGGCUCUAAACCCCGC GACCGGGGUUCGAAUCCCCGCGGGGCUACCA
SEQ ID: 66	AE(NN) tRNA library	GCCCCGGUGNUGUAGCCCGGCCAAGCANGCGGGCCUCUAAAGCCC GCGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 67	AE (GC) tRNA	GCCCCGGUGGUGUAGCCCGGCCAAGCACGCGGGCCUCUAAAGCCC GCGACCCGGGUUCAAAUCCCGGCCGGGGCACCA
SEQ ID: 68	AfERS	gtgaaagaag tcataatgaa atacgtagtt cagaacgccg caaagtacgg aaaagccagt gaaaaggctg taatgggggaa ggtgatggca gaaaatccag aactcagaaa aaaagctaaa gaggtccttg aactcgtaaa ggagtgcatt accgagttcg aagcgctttc tgaagaagta aggaaggagc taatcaaaaa atacagcatg gatagcgaag ctaaaaggga gttggagacg aagaagcttc cagagcttga gggggctgag aaagggaaag ttgtgatgag attcgctcct aatccaaacg gtccacctac actcggttct gctaggggga taatcgtcaa cggtgaatac gcgaagatgt acgaagggaa gtacattatc aggtttgacg atacagaccc cagaaccaag agaccgatga ttgaggccta cgagtggtac ctagaggaca ttgaatggct cggttacaag cctgacgagg ttatttacgc ctcaagaagg attcccatct actacgatta tgcaagaaag cttatagaaa tgggcaaagc ctacacctgc ttctgcagtc aggaggagtt caagaagttc agggacagtg gtgaggagtg cccacacaga aacatcagcg tcgaggatac acttgaagtc tgggagagaa tgcttgaagg agactatgag gaaggagaag tcgttcttag aattaagacg gatatgcgcc acaaggatcc ggcaataagg gactgggtgg ctttcaggat aataaaggaa tcacatcctc tcgttggggga taaatatgtt gtctatccca cactcgattt tgaatctgct atagaggatc accttttagg cataacgcac attatcaggg gtaaggactt aattgattct gagaggaggc agagatacat ctacgagtat tttggctgga tttaccccat tacaaagcac tggggcaggg ttaaaatctt cgaattcggg aagctatcga cttcttcaat taagaaggat attgaaaggg gtaagtatga gggctgggat gacccaaggc tgccaacctt gagagccttc aggaggagag gatttgagcc tgaagccata aagagcttct tcctttcgct gggagttggc gagaacgacg tttccgtcag tctcaagaat ctttacgctg agaacaggaa aatcatcgat cgcaaggcaa accgttactt tttcatttgg gggcccgtga agattgaaat cgttaacctg ccggagaaga aagaggtcga gctcccgctg aatccgcaca cgggcgagaa gagaaggctg aaaggtgaaa gaactatata tgttacaaaa gacgacttcg agaggttgaa agggcaggtt gtaaggctga aggacttctg caacgttctg cttgacgaga aggcagagtt catgggtttt gagcttgagg gggtcaagaa ggggaagaac ataatccact ggcttcctga gagcgaggct atcaagggca aggttatcgg cgaaagggaa gcggaagggc ttgtggagag aaacgcagtt agagatgtcg ggaaagtcgt gcagttcgag agatttgctt tctgtaaagt agagtctgct gacgaggaac tggtcgcggt gtacacgcac ccgtga
SEQ ID: 69	AfERS.pep	VKEVIMKYVV QNAAKYGKAS EKAVMGKVMA ENPELRKKAK EVLELVKECI TEFEALSEEV RKELIKKYSM DSEAKRELET KKLPELEGAE KGKVVMRFAP NPNGPPTLGS ARGIIVNGEY AKMYEGKYII RFDDTDPRTK RPMIEAYEWY LEDIEWLGYK PDEVIYASRR IPIYYDYARK LIEMGKAYTC FCSQEEFKKF RDSGEECPHR NISVEDTLEV WERMLEGDYE EGEVVLRIKT DMRHKDPAIR DWVAFRIIKE SHPLVGDKYV VYPTLDFESA IEDHLLGITH IIRGKDLIDS ERRQRYIYEY FGWIYPITKH WGRVKIFEFG KLSTSSIKKD IERGKYEGWD DPRLPTLRAF RRRGFEPEAI KSFFLSLGVG ENDVSVSLKN LYAENRKIID RKANRYFFIW GPVKIEIVNL PEKKEVELPL NPHTGEKRRL KGERTIYVTK DDFERLKGQV VRLKDFCNVL LDEKAEFMGF ELEGVKKGKN IIHWLPESEA IKGKVIGERE AEGLVERNAV RDVGKVVQFE RFAFCKVESA DEELVAVYTH P*
SEQ ID: 70	ApERS	atgtcaatgc tcctagggga ccatcctgag cttaggagta gggcgaggga gatagcctcc ctggctgcaa gggtcgtgga gcaggtgaac agcatgcctg

		ctggcgagca	gaagaggctc	ctctcagaac	agtatccgga	gctagcgagg
		tttgaagagc	agcgcgagaa	aggcgataaa	ggcctacccc	cgctgcccgg
		ggccgtggaa	ggcagggtca	agctgaggtt	cgcccccaac	ccagactttg
		tgatacacat	gggcaacgct	aggccggcca	tagtgaacca	cgagtacgcc
		aggatgtata	aggggaggat	ggtgctgcgg	ttcgaagaca	cagaccctag
		gacaaagact	cccctacggg	aggcgtacga	cctcatccgc	caggacctca
		agtggcttgg	cgtttcctgg	gacgaggagt	acatccagag	cctcaggatg
		gaggtcttct	acagcgtagc	cagaagggct	atagagcggg	ggtgcgccta
		cgtcgacaac	tgtgggagag	agggtaagga	gctgctctcg	cgtggagagt
		attgtccgac	cagggatctt	ggccccgagg	ataatctgga	gctgtttgag
		aagatgcttg	aaggagagtt	ttacgagggg	gaggctgttg	tgaggatgaa
		gacggacccg	cgccacccca	accccagcct	gagggactgg	gtggccatga
		ggattataga	cacggagaaa	cacccccacc	cctcgtgggg	gagcaggtac
		ctggtgtggc	cgacctataa	cttcgccgtc	agcgtcgacg	accacatgat
		ggagataacc	cacgtgctca	ggggtaagga	gcaccagctg	aacaccgaga
		agcagcttgc	cgtctacagg	tgtatgggct	ggaggccgcc	gtacttcatc
		cacttcggca	ggctaaagct	ggaggggttt	atactgagca	agagcaagat
		aagaaagctc	ctcgaagaga	ggccaggaga	gttcatgggc	tacgacgacc
		cccgcttcgg	cacgatagcc	gggctgagga	ggaggggtgt	gctggcggag
		gccataagac	agataatact	cgaggtgggg	gtcaagccca	cggacgctac
		aataagctgg	gcaaacctgg	cggccgctaa	taggaagctg	ctggacgagc
		gggcggacag	gataatgtat	gtcgaagacc	ccgttgagat	ggaggtcgag
		ctggcccagg	tggagtgcag	agccgcggag	ataccgttcc	acccgagccg
		cccccagagg	aagaggagga	taaccttgtg	cacgggggac	aaggttctcc
		tcacaaggga	ggacgcagta	gagggtaggc	aactgaggct	catgggccta
		tcaaacttca	ccgtcagcca	aggcatattg	agggaggtgg	atcccagcct
		cgagtatgcg	aggagaatga	agctacctat	agtacaatgg	gtgaagaagg
		gcggagaagc	tagcgtagag	gtactcgagc	ccgtcgagct	ggagctcagg
		aggcaccagg	gctacgctga	ggacgctatt	aggggctatg	gggttgacag
		ccgcctacag	tttgtaaggt	acggttttgt	gagggtggac	agcgtggaag
		acggggtcta	cagggtgata	tacacacaca	agtag	

SEQ ID: 71	ApERS.pep	MSMLLGDHPE LRSRAREIAS LAARVVEQVN SMPAGEQKRL LSEQYPELAR FEEQREKGDK GLPPLPGAVE GRVKLRFAPN PDFVIHMGNA RPAIVNHEYA RMYKGRMVLR FEDTDPRTKT PLREAYDLIR QDLKWLGVSW DEEYIQSLRM EVFYSVARRA IERGCAYVDN CGREGKELLS RGEYCPTRDL GPEDNLELFE KMLEGEFYEG EAVVRMKTDP RHPNPSLRDW VAMRIIDTEK HPHPLVGSRY LVWPTYNFAV SVDDHMMEIT HVLRGKEHQL NTEKQLAVYR CMGWRPPYFI HFGRLKLEGF ILSKSKIRKL LEERPGEFMG YDDPRFGTIA GLRRRGVLAE AIRQIILEVG VKPTDATISW ANLAAANRKL LDERADRIMY VEDPVEMEVE LAQVECRAAE IPFHPSRPQR KRRITLCTGD KVLLTREDAV EGRQLRLMGL SNFTVSQGIL REVDPSLEYA RRMKLPIVQW VKKGGEASVE VLEPVELELR RHQGYAEDAI RGYGVDSRLQ FVRYGFVRVD SVEDGVYRVI YTHK*

SEQ ID: 72	MmERS	atgaccttaa gtcctgaaga tattaaaata attgaaaaat acgctcttca aaatgctgta aaatacggaa aagccccaca gcccaaagct gttatgggta aagttatggg ggaatgtccc cagttgaggg ctgaccccag cggggtatct gctgcccttg aaaatatagt ttcagaaatt gctaaaggaa accctgaagc ctgggaagcc cgcctgtcgg aaatcgctcc cgagcttata gaggctctga gtgtaaagaa agagcctgac aaaggcttaa aacctcttga gggagccgaa accggtaagg ttgtaatgcg ctttgcccca aatcccaacg acccgcaac cctgggaagt tcaaggggaa tggttgtcaa ttccgagtat gtaaaagtgt ataaaggaaa gtttatcctg cgctttgatg acactgaccc tgatataaag cgtcccatgc tcgaagccta tgactggtac atggacgact ttaaatggct cggggttgtg cctgaccagg ttgtacgtgc ctctgaccgt ttcccgattt attacgatta tgcaagaaaa ctgattgaga tgggaaaagc ctatgtctgt ttctgtaaag gagaagactt caaaaggtta aaagactcca aaaaggcctg ccctcacagg gacacagacc ctgaagaaaa ccttatgcac tgggaaaaga tgcttgccgg agaatatgaa gaccagcagg ctgttctgcg cataaagacc

		gatatagagc	acaaagaccc	tgcacttcgg	gactggggag	cattcaggat
		tagaaaaatg	tcccacccac	gccctgagat	tggaaataaa	tacatcgtct
		ggcccccttct	ggactttgca	ggtgctatcg	aagaccacga	gctcggtatg
		acccacatca	tccggggtaa	agaccttatc	gacagtgaaa	aacggcagac
		atacatttac	aaatacttcg	gctggacata	cccgaaaaca	actcactggg
		gcagggtaaa	gatccatgag	ttcggaaagt	tcagcacaag	taccctgcgt
		aaagctatcg	aatccggaga	atacagcggc	tgggacgacc	cccgcctgcc
		gacaataagg	gcaatcaggc	gccgtggaat	tagggctgaa	gctctcaaaa
		agttcatgat	agaaatggga	gtcgggatga	ccgatgtgag	catcagcatg
		gaatctcttt	atgccgagaa	ccgcaaaatc	gtggacccgg	ttgcaaaccg
		ctatttcttt	gtctggaacc	ctgtggagct	tgaaattgaa	ggtatgaaac
		ctgtggtcgc	aaaggtcccc	cgtcacccaa	cagaccatgc	aagaggaatg
		agggaaatct	ctatcgaaaa	taaggttctt	gtctgtgccg	aggacattga
		aaagctcaat	gtaggttctg	tcctccgctt	aaaagatctc	tgcaacgttg
		aaattacttc	cctttccccg	ctgcgggtta	agcgctccga	aacctctctc
		gaagacctca	aaaaagcaaa	aggaaaaatc	atccactggg	tgccggttga
		tggaattcct	gtgaaagtct	gcgggccaga	gggcgatatc	gaagggacag
		gagaacgcgg	gatcgagacc	gagcttgata	atattgtcca	gtttgaacgc
		ttcggtttct	gcaggatcga	cgccgtagac	ggagaaaatg	tcgtggcgta
		ctttgctcac	aaatga			
SEQ ID: 73	MmERS.pep	MTLSPEDIKI	IEKYALQNAV	KYGKAPQPKA	VMGKVMGECP	QLRADPSGVS
		AALENIVSEI	AKGNPEAWEA	RLSEIAPELI	EALSVKKEPD	KGLKPLEGAE
		TGKVVMRFAP	NPNGPATLGS	SRGMVVNSEY	VKVYKGKFIL	RFDDTDPDIK
		RPMLEAYDWY	MDDFKWLGVV	PDQVVRASDR	FPIYYDYARK	LIEMGKAYVC
		FCKGEDFKRL	KDSKKACPHR	DTDPEENLMH	WEKMLAGEYE	DQQAVLRIKT
		DIEHKDPALR	DWGAFRIRKM	SHPRPEIGNK	YIVWPLLDFA	GAIEDHELGM
		THIIRGKDLI	DSEKRQTYIY	KYFGWTYPKT	THWGRVKIHE	FGKFSTSTLR
		KAIESGEYSG	WDDPRLPTIR	AIRRRGIRAE	ALKKFMIEMG	VGMTDVSISM
		ESLYAENRKI	VDPVANRYFF	VWNPVELEIE	GMKPVVAKVP	RHPTDHARGM
		REISIENKVL	VCAEDIEKLN	VGSVLRLKDL	CNVEITSLSP	LRVKRSETSL
		EDLKKAKGKI	IHWVPVDGIP	VKVCGPEGDI	EGTGERGIET	ELDNIVQFER
		FGFCRIDAVD	GENVVAYFAH	K*		
SEQ ID: 74	MtERS	gtggtgccag	tggaggacct	tgtctacagg	tacgcactcc	tgaacgctgt
		gaaacacagg	ggaagggcaa	acccggggggc	cgtcatgggg	gccgtcatga
		gcaacgaacc	tgaactcagg	aagatggcgc	cccaggttaa	ggaagcagtt
		gaggccgctg	ttgagagggt	taacagttta	agccccgagg	agcagcagca
		ggagatggag	aggctgggcc	ttgagatcac	cgagaggaaa	cagaagaagc
		ggaagggtct	aagggaactg	gccggggtga	agggtgaggt	ggtcctcagg
		ttcgcccccca	accccagcgg	acccctccac	ataggccatg	caagggccgc
		gatcctcaac	catgaatatg	caaggaaata	tgacggcagg	ctcatcctca
		ggatagagga	cacggacccc	cgcagggttg	acccggaggc	ctacgatatg
		attccagccg	accttgagtg	gctgggcgtt	gaatgggatg	agacagttat
		ccagagcgac	cgcatggaaa	cctactatga	gtacacagag	aaactcatag
		agaggggtgg	tgcgtatgta	tgcacatgca	ggccggagga	gttcagggaa
		ctcaagaaca	ggggagaggc	ctgtcactgc	aggtcccttg	gtttcaggga
		gaacctccag	cgatggaggg	aaatgttcga	gatgaaggag	ggctcagccg
		ttgtgagggt	taaaacggac	ctgaaccacc	caaaccctgc	cataagagac
		tgggtctcaa	tgaggatcgt	tgaggcagag	cacccacgca	ccggtacacg
		ctacagggtc	taccccatga	tgaacttctc	agtggcggtt	gatgaccacc
		tccttggcgt	gacacacgtc	ctgaggggta	aggaccacct	ggcaaacaga
		gagaagcagg	agtacctcta	caggcacctt	ggctgggagc	ccccgaatt
		catacactac	gggcgcctga	agatggacga	cgttgcactc	agcacctcgg
		gggcccgtga	gggcatcctc	aggggtgagt	attctggatg	ggacgaccca
		cgcctcggaa	ccctgagggc	cattgcaagg	aggggtatac	gaccggaggc
		cataagaaag	ttaatggttg	aaatcggcgt	aaagatagca	gattccacaa
		tgagctggaa	gaagatctac	ggcctcaaca	ggagcatcct	cgaagaggag
		gccaggaggt	acttctttgc	tgcagatcct	gttaaacttg	aagtggtggg
		tttacccgga	cctgtcaggg	ttgaaaggcc	cctacacccg	gaccaccctg

		aaatcgggaa	cagggtactt	gaactgaggg	gtgaggtgta cctgccgggc
		gatgaccttg	gggaaggacc	gctgcgcctc	atagacgcgg ttaacgtgat
		ctactctggg	ggtgaactga	ggtaccacag	tgagggcatc gaggaggccc
		gtgaactcgg	cgcttcaatg	atacactggg	tccctgcaga gtccgccctt
		gaggccgagg	tcatcatgcc	ggacgcctcc	agggtgaggg gagtcatcga
		ggctgacgca	tcagaactgg	aggttgatga	tgttgtgcag cttgagaggt
		tcggcttcgc	ccgcctggat	tctgcaggcc	caggaatggt cttctactat
		gcccacaaat	ga		

SEQ ID: 75	MtERS.pep	VVPVEDLVYR	YALLNAVKHR	GRANPGAVMG	AVMSNEPELR KMAPQVKEAV
		EAAVERVNSL	SPEEQQQEME	RLGLEITERK	QKKRKGLREL AGVKGEVVLR
		FAPNPSGPLH	IGHARAAILN	HEYARKYDGR	LILRIEDTDP RRVDPEAYDM
		IPADLEWLGV	EWDETVIQSD	RMETYYEYTE	KLIERGGAYV CTCRPEEFRE
		LKNRGEACHC	RSLGFRENLQ	RWREMFEMKE	GSAVVRVKTD LNHPNPAIRD
		WVSMRIVEAE	HPRTGTRYRV	YPMMNFSVAV	DDHLLGVTHV LRGKDHLANR
		EKQEYLYRHL	GWEPPEFIHY	GRLKMDDVAL	STSGAREGIL RGEYSGWDDP
		RLGTLRAIAR	RGIRPEAIRK	LMVEIGVKIA	DSTMSWKKIY GLNRSILEEE
		ARRYFFAADP	VKLEVVGLPG	PVRVERPLHP	DHPEIGNRVL ELRGEVYLPG
		DDLGEGPLRL	IDAVNVIYSG	GELRYHSEGI	EEARELGASM IHWVPAESAL
		EAEVIMPDAS	RVRGVIEADA	SELEVDDVVQ	LERFGFARLD SAGPGMVFYY
		AHK*			

SEQ ID: 76	PhERS	atggatgtgg	aaaagatagc	tctcaaacat	gcccttatta atgcaattga
		gcatggaggt	aaggctaatc	ttaaggccgt	catcggtaaa gtgctcggcg
		agaaccctga	gcttaggccc	agggccaagg	agataattcc cataattaat
		aaggtagtgg	aagaggtcaa	ctcactagca	agggatgagc agctggaaaa
		gcttaaggat	atatatccag	agtactttga	aaagaaggaa gagaagaagg
		aaaagaaggg	actacctctt	cttcctaaag	cagagaaagg caaggtcgtt
		actcgctttg	ctccaaatcc	agatggtgct	tttcacctgg gtaatgcaag
		ggccgcaata	cttagttatg	aatacgccaa	gatgtacggt gggaagttta
		tactccgctt	tgatgatacc	gatccaaaag	tcaagaggcc tgaacccata
		ttctacaaga	tgataattga	agatctagag	tggcttggaa taaagccaga
		tgaaatagtg	tacgctagtg	atagactcga	aatttactac aagtacgctg
		aggaactaat	aaaaatgggt	aaagcatatg	tctgcacttg tcctccagag
		aagttcagag	agcttagaga	taagggaatt	ccatgtccgc atagagatga
		gcctgtggaa	gttcagctcg	agcgctggaa	gaagatgtta aacggcgaat
		ataaagaagg	agaagctgtt	gtaaggataa	agactgactt aaatcaccca
		aatcctgcgg	ttagggattg	gcccgcactt	agaataattg ataacccaaa
		ccatccgaga	accggaaata	agtacagggt	atggcctctc tacaattttg
		cctcagcgat	tgatgatcac	gagcttggag	ttacccatat cttccgtgga
		caggagcacg	ccgagaacga	gactagacag	aggtatatct acgagtactt
		tggttgggag	tatcccgtca	cgatccatca	cgggaggctt agcatagagg
		gagttgtttt	aagcaaatcc	aagacaagga	aaggaataga agaaggtaag
		tatctcggtt	gggatgatcc	taggcttgga	acgataagag ctttgaggag
		gaggggaatt	cttcctgaag	ccataaaaga	attgatcata gaagttggat
		tgaagaaaag	tgacgcgacg	atcagttggg	agaatctggc agcgataaac
		agaaaacttg	ttgatccaat	tgccaataga	tatttctttg ttgcagatcc
		aattccgatg	gaagtcgaag	gggctcctga	atttatagca gagataccgc
		ttcatcctga	ccatcctgaa	aggggcgtta	gaaggcttaa gttcactcca
		gaaagacctg	tttacgtttc	gaaagacgat	ttgaatctct aaaaccagg
		caattttgtg	aggttaaagg	atctcttcaa	tgttgaaata cttgaagttg
		gagacaaaat	aagagctagg	ttctacagct	ttgaatatga aattgcaaag
		aagaacaggt	ggaagatggt	tcactgggtt	accgaaggta gaccctgtga
		ggttataatt	cctgaaggag	acgagcttgt	agtgaggaag ggattgctcg
		agaggatgc	caaagttcaa	gttaacgaaa	tagtccagtt tgaacgtttc
		ggttttgtaa	ggatagacag	aattgagggg	gataaagtta tcgcaatata
		tgctcacaag	taa		

SEQ ID: 77	PhERS.pep	MDVEKIALKH	ALINAIEHGG	KANLKAVIGK	VLGENPELRP RAKEIIPIIN

		KVVEEVNSLA	RDEQLEKLKD	IYPEYFEKKE	EKKEKKGLPL	LPKAEKGKVV
		TRFAPNPDGA	FHLGNARAAI	LSYEYAKMYG	GKFILRFDDT	DPKVKRPEPI
		FYKMIIEDLE	WLGIKPDEIV	YASDRLEIYY	KYAEELIKMG	KAYVCTCPPE
		KFRELRDKGI	PCPHRDEPVE	VQLERWKKML	NGEYKEGEAV	VRIKTDLNHP
		NPAVRDWPAL	RIIDNPNHPR	TGNKYRVWPL	YNFASAIDDH	ELGVTHIFRG
		QEHAENETRQ	RYIYEYFGWE	YPVTIHHGRL	SIEGVVLSKS	KTRKGIEEGK
		YLGWDDPRLG	TIRALRRRGI	LPEAIKELII	EVGLKKSDAT	ISWENLAAIN
		RKLVDPIANR	YFFVADPIPM	EVEGAPEFIA	EIPLHPDHPE	RGVRRLKFTP
		ERPVYVSKDD	LNLLKPGNFV	RLKDLFNVEI	LEVGDKIRAR	FYSFEYEIAK
		KNRWKMVHWV	TEGRPCEVII	PEGDELVVRK	GLLEKDAKVQ	VNEIVQFERF
		GFVRIDRIEG	DKVIAIYAHK	*		

SEQ ID: 78	SsERS	ttgatggaat	taaatgaatt	acgagaaact	atttataaat	atgctcttca
		aaatgcggta	aagcataatg	ggaaagccga	aacgggtcca	gtgatgagta
		agattatagc	ggagaggccg	gaattaagat	caaatgctag	agaaattgta
		aagattatca	aggagatcat	agaacaagtt	aactcattaa	ctttagagca
		gcagttaact	gaaattaaag	caaagtatcc	agagttatta	gaagagaaga
		agcatgaaga	aaaaagaaaa	gtgttgccac	ctcttcctaa	tgttaaagga
		caagtagtaa	ctagatttgc	ccctaatccg	gatggtcctc	ttcatttagg
		taacgcaagg	tctgcgattc	tatcttacga	atacgctaaa	atgtataacg
		gtaagttcat	actgaggttt	gatgatacag	atcctaaggt	taaaaggcca
		attctagatg	catatgattg	gattaaggag	gatctaaaat	ggctgggaat
		taaatgggaa	caagaacttt	acgcttcgga	aaggctcgag	ttatactata
		agtacgctcg	gtatttgata	gaaaaaggat	atgcgtatgt	tgatacatgt
		gatagtagta	ttttcaggaa	gtttagggat	agtagaggta	aaatgaaaga
		gccagaatgt	ttacatagga	gctcttcacc	agagagtaat	ctggagctgt
		ttgaaaaaat	gctaggagga	aagtttaaag	aaggagaagc	tgtggttagg
		ttaaagaccg	atctatcgga	tccagatcca	tctcaaatag	attgggtaat
		gcttagaata	attgatactg	ctaaaaatcc	tcatccacga	gtagggagca
		agtattgggt	atggcctacc	tataactttg	cttctataat	agatgatcat
		gaacttggaa	ttactcatgt	actgagagct	aaggagcata	tgtcaaatac
		agaaaagcag	agatacattt	ctgagtacat	gggatgggaa	tttccagagg
		tcttacagtt	tggaagacta	agacttgaag	gttttatgat	gagtaagtcg
		aaaataagag	gaatgttaga	gaagggtacc	aatagggatg	atcctagatt
		gccaacttta	gctggactca	gaagaagagg	aatcttacca	gatacaatta
		aagacgtaat	aattgatgtc	ggagtaaagg	taactgatgc	tactataagt
		tttgagaaca	ttgccgcaat	taataggaaa	aaactagatc	cggtagctaa
		aagaataatg	tttgtaaaag	atgcagaaga	atttagcgtc	gagctacctg
		aatcactaaa	tgctaaaata	cctttaattc	cctctaaaca	agaaatgaat
		agaactatta	ttgtcaatcc	aggagataaa	attttgatag	aatctaacga
		tgcagaagat	aatagtatac	taaggttaat	ggaattatgc	aatgttaaag
		ttgataagca	taatcgtaag	ctcatatttc	acagcaaaac	gttagacgag
		gctaagaaag	tcaatgcaaa	gatagttcaa	tgggtaaaat	caaatgaaaa
		agtaccagta	atggtagaga	aagcggaaag	agatgagata	aaaatgataa
		atggttatgc	tgaaaaaata	gcagcagatt	tggagataga	cgaaattgta
		caattttacc	gatttggatt	tgtaagagta	gataggaaag	atgagaatat
		gctacgtgtt	gtattctcac	acgactga		

SEQ ID: 79	SsERS.pep	LMELNELRET	IYKYALQNAV	KHNGKAETGP	VMSKIIAERP	ELRSNAREIV
		KIIKEIIEQV	NSLTLEQQLT	EIKAKYPELL	EEKKHEEKRK	VLPPLPNVKG
		QVVTRFAPNP	DGPLHLGNAR	SAILSYEYAK	MYNGKFILRF	DDTDPKVKRP
		ILDAYDWIKE	DLKWLGIKWE	QELYASERLE	LYYKYARYLI	EKGYAYVDTC
		DSSIFRKFRD	SRGKMKEPEC	LHRSSSPESN	LELFEKMLGG	KFKEGEAVVR
		LKTDLSDPDP	SQIDWVMLRI	IDTAKNPHPR	VGSKYWVWPT	YNFASIIDDH
		ELGITHVLRA	KEHMSNTEKQ	RYISEYMGWE	FPEVLQFGRL	RLEGFMMSKS
		KIRGMLEKGT	NRDDPRLPTL	AGLRRRGILP	DTIKDVIIDV	GVKVTDATIS
		FENIAAINRK	KLDPVAKRIM	FVKDAEEFSV	ELPESLNAKI	PLIPSKQEMN
		RTIIVNPGDK	ILIESNDAED	NSILRLMELC	NVKVDKHNRK	LIFHSKTLDE
		AKKVNAKIVQ	WVKSNEKVPV	MVEKAERDEI	KMINGYAEKI	AADLEIDEIV

		QFYRFGFVRV DRKDENMLRV VFSHD*			
SEQ ID: 80	pACKO-A184TAG	gaactccgga	tgagcattca	tcaggcgggc	aagaatgtga	ataaaggccg
		gataaaactt	gtgcttattt	ttctttacgg	tctttaaaaa	ggccgtaata
		tccagctgaa	cggtctggtt	ataggtacat	tgagcaactg	actgaaatgc
		ctcaaaatgt	tctttacgat	gccattggga	tatatcaacg	gtggtatatc
		cagtgatttt	tttctccatt	ttagcttcct	tagctcctga	aaatctcgat
		aactcaaaaa	atacgcccgg	tagtgatctt	atttcattat	ggtgaaagtt
		ggaacctctt	acgtgccgat	caacgtctca	ttttcgccaa	aagttggccc
		agggcttccc	ggtatcaaca	gggacaccag	gatttattta	ttctgcgaag
		tgatcttccg	tcacaggtat	ttattcggcg	caaagtgcgt	cgggtgatgc
		tgccaactta	ctgatttagt	gtatgatggt	gtttttgagg	tgctccagtg
		gcttctgttt	ctatcagctg	tccctcctgt	tcagctactg	acggggtggt
		gcgtaacggc	aaaagcaccg	ccggacatca	gcgctagcgg	agtgtatact
		ggcttactat	gttggcactg	atgagggtgt	cagtgaagtg	cttcatgtgg
		caggagaaaa	aaggctgcac	cggtgcgtca	gcagaatatg	tgatacagga
		tatattccgc	ttcctcgctc	actgactcgc	tacgctcggt	cgttcgactg
		cggcgagcgg	aaatggctta	cgaacggggc	ggagatttcc	tggaagatgc
		caggaagata	cttaacaggg	aagtgagagg	gccgcggcaa	agccgttttt
		ccataggctc	cgccccctg	acaagcatca	cgaaatctga	cgctcaaatc
		agtggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct
		ggcggctccc	tcgtgcgctc	tcctgttcct	gcctttcggt	ttaccggtgt
		cattccgctg	ttatggccgc	gtttgtctca	ttccacgcct	gacactcagt
		tccgggtagg	cagttcgctc	caagctggac	tgtatgcacg	aacccccgt
		tcagtccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc
		cggaaagaca	tgcaaaagca	ccactggcag	cagccactgg	taattgattt
		agaggagtta	gtcttgaagt	catgcgccgg	ttaaggctaa	actgaaagga
		caagttttgg	tgactgcgct	cctccaagcc	agttacctcg	gttcaaagag
		ttggtagctc	agagaacctt	cgaaaaaccg	ccctgcaagg	cggttttttc
		gttttcagag	caagagatta	cgcgcagacc	aaaacgatct	caagaagatc
		atcttattaa	tcagataaaa	tatttctaga	tttcagtgca	atttatctct
		tcaaatgtag	cacctgaagt	cagccccata	cgatataagt	tgtaattctc
		atgtttgaca	gcttatcatc	gataagcttt	aatgcggtag	tttatcacag
		ttaaattgct	aacgcagtca	ggcaccgtgt	atgaaatcta	acaatgcgct
		catcgtcatc	ctcggcaccg	tcaccctgga	tgctgtaggc	ataggcttgg
		ttatgccggt	actgccgggc	ctcttgcggg	atatcGGTTT	CTTAGACGTC
		AGGTGGCACT	TTtcggggaa	atgtgcgcgg	aacccctatt	tgtttatttt
		tctaaataca	ttcaaatatg	tatccgctca	tgagacaata	accctgataa
		atgcttcaat	aatattgaaa	aaggaagagt	atgagtattc	aacatttccg
		tgtcgccctt	attccctttt	ttgcggcatt	ttgccttcct	gtttttgctc
		acccagaaac	gctggtgaaa	gtaaaagatg	ctgaagatca	gttgggtgca
		cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag
		ttttcgcccc	gaagaacgtt	ttccaatgat	gagcactttt	aaagttctgc
		tatgtggcgc	ggtattatcc	cgtgttgacg	ccgggcaaga	gcaactcggt
		cgccgcatac	actattctca	gaatgacttg	gttgagtact	caccagtcac
		agaaaagcat	cttacggatg	gcatgacagt	aagagaatta	tgcagtgctg
		ccataaccat	gagtgataac	actgcggcca	acttacttct	gacaacgatc
		ggaggaccga	aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt
		aactcgcctt	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaacg
		acgagcgtga	caccacgatg	cctTAGgcaa	tggcaacaac	gttgcgcaaa
		ctattaactg	gcgaactact	tactctagct	tcccggcaac	aattaataga
		ctggatggag	gcggataaag	ttgcaggacc	acttctgcgc	tcggcccttc
		cggctggctg	gtttattgct	gataaatctg	gagccggtga	gcgtgggtct
		cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	cccgtatcgt
		agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac
		agatcgctga	gataggtgcc	tcactgatta	agcattggca	ccaccaccac
		caccactaaC	CCGGGACCAA	GTTTACTCAT	ATATACttta	gattgattta
		aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa
		tCTCATGACC	AAAATCCCTT	AACGgcatgc	accattcctt	gcggcggcgg

		tgctcaacgg cctcaaccta ctactGGGCT GCTTCCTAAT GCAGGAGTCG CATAAGGGAG AGCGTCTGGC GAAAGGGGGA TGTGCTGCAA GGCGATTAAG TTGGGTAACG CCAGGGTTTT CCCAGTCACG ACGTTGTAAA ACGACGGCCA GTGCCAAGCT TAAAAAaaat ccttagcttt cgctaaggat CTGCAGTTAT AATCTCTTTC TAATTGGCTC TAAAATCTTT ATAAGTTCTT CAGCTACAGC ATTTTTTAAA TCCATTGGAT GCAATTCCTT ATTTTTAAAT AAACTCTCTA ACTCCTCATA GCTATTAACT GTCAAATCTC CACCAAATTT TTCTGGCCTT TTTATGGTTA AAGGATATTC AAGGAAGTAT TTAGCTATCT CCATTATTGG ATTTCCTTCA ACAACTCCAG CTGGGCAGTA TGCTTTCTTT ATCTTAGCCC TAATCTCTTC TGGAGAGTCA TCAACAGCTA TAAAATTCCC TTTTGAAGAA CTCATCTTTC CTTCTCCATC CAAACCCGTT AAGACAGGGT TGTGAATACA AACAACCTTT TTTGGTAAAA GCTCCCTTGC TAACATGTGT ATTTTTCTCT GCTCCATCCC TCCAACTGCA ACATCAACGC CTAAATAATG AATATCATTA ACCTGCATTA TTGGATAGAT AACTTCAGCA ACCTTTGGAT TTTCATCCTC TCTTGCTATA AGTTCCATAC TCCTTCTTGC TCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATAGACATTC AGTGTATAAT CCTTATCAAG CTGGAATTCa gcgttacaag tattacacaa agtttttat gttgagaata ttttttgat ggggcgccac ttatttttga tcgttcgctc aaagAAGCGG CGCCAGGGNT GTTTTTCTTT TCACCAGTNA GACGGGCAAC AGAACGCCAT Gagcggcctc atttcttatt ctgagttaca acagtccgca ccgctgtccg gtagctcctt ccggtgggcg cggggcatga ctatcgtcgc cgcacttatg actgtcttct ttatcatgca actcgtagga caggtgccgg cagcgcccaa cagtcccccg gccacggggc ctgccaccat acccacgccg aaacaagcgc cctgcaccat tatgttccgg atctgcatcg caggatgctg ctggctaccc tgtggaacac ctacatctgt attaacgaag cgctaaccgt ttttatcagg ctctgggagg cagaataaat gatcatatcg tcaattatta cctccacggg gagagcctga gcaaactggc ctcaggcatt tgagaagcac acggtcacac tgcttccggt agtcaataaa ccggtaaacc agcaatagac ataagcggct atttaacgac cctgccctga accgacgacc gggtcgaatt tgctttcgaa tttctgccat tcatccgctt attatcactt attcaggcgt agcaccaggc gtttaagggc accaataact gccttaaaaa aattacgccc cgccctgcca ctcatcgcag tactgttgta attcattaag cattctgccg acatggaagc catcacagac ggcatgatga acctgaatcg ccagcggcat cagcaccttg tcgccttgcg tataatattt gcccatggtg aaaacggggg cgaagaagtt gtccatattg gccacgttta aatcaaaact ggtgaaactc acccagggat tggctgagac gaaaaacata ttctcaataa acccttagg gaaataggcc aggttttcac cgtaacacgc cacatcttgc gaatatatgt gtagaaactg ccggaaatcg tcgtggtatt cactccagag cgatgaaaac gtttcagttt gctcatggaa aacggtgtaa caagggtgaa cactatccca tatcaccagc tcaccgtctt tcattgccat acg
SEQ ID: 81	pACKO- Bla	gaactccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttatta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaagatgc caggaagata cttaacaggg aagtgagagg gccgcggcaa agccgttttt ccataggctc cgcccccctg acaagcatca cgaaatctga cgctcaaatc

```
agtggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct
ggcggctccc tcgtgcgctc tcctgttcct gcctttcggt ttaccggtgt
cattccgctg ttatggccgc gtttgtctca ttccacgcct gacactcagt
tccgggtagg cagttcgctc caagctggac tgtatgcacg aacccccgt
tcagtccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc
cggaaagaca tgcaaaagca ccactggcag cagccactgg taattgattt
agaggagtta gtcttgaagt catgcgccgg ttaaggctaa actgaaagga
caagttttgg tgactgcgct cctccaagcc agttacctcg gttcaaagag
ttggtagctc agagaacctt cgaaaaaccg ccctgcaagg cggttttttc
gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc
atcttattaa tcagataaaa tatttctaga tttcagtgca atttatctct
tcaaatgtag cacctgaagt cagccccata cgatataagt tgtaattctc
atgtttgaca gcttatcatc gataagcttt aatgcggtag tttatcacag
ttaaattgct aacgcagtca ggcaccgtgt atgaaatcta acaatgcgct
catcgtcatc ctcggcaccg tcaccctgga tgctgtaggc ataggcttgg
ttatgccggt actgccgggc ctcttgcggg atatcGGTTT CTTAGACGTC
AGGTGGCact tttcgggga atgtgcgcgg aaccctatt tgtttatttt
tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa
atgcttcaat aatattgaaa aaggaagagt atgagtattc aacatttccg
tgtcgccctt attcccttt ttgcggcatt ttgccttCCT GTTTTTGCTC
ACCCAGAAAC ACTAGtgcag caatggcaac aacgttgcgc aaactattaa
ctggcgaact acttactcta gcttcccggc aacaattaat agactggatg
gaggcggata aagttgcagg accacttctg cgctcggccc ttccggctgg
ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta
tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc
tacacgacgg ggagtcaggc aactatggat gaacgaaata gacagatcgc
tgagataggt gcctCACTGA TTAAGCATTG GTAACCCGGG ACCAAGTTTA
CTCATATATA Ctttagattg atttaaaact tcattttaa tttaaaagga
tctaggtgaa gatcctttt gataatCTCA TGACCAAAAT CCCTTAACGg
catgcaccat tccttgcggc ggcggtgctc aacggcctca acctactact
GGGCTGCTTC CTAATGCAGG AGTCGCATAA GGGAGAGCGT CTGGCGAAAG
GGGGATGTGC TGCAAGGCGA TTAAGTTGGG TAACGCCAGG GTTTTCCCAG
TCACGACGTT GTAAAACGAC GGCCAGTGCC AAGCTTAAAA Aaaatcctta
gctttcgcta aggatCTGCA GTTATAATCT CTTTCTAATT GGCTCTAAAA
TCTTTATAAG TTCTTCAGCT ACAGCATTTT TTAAATCCAT TGGATGCAAT
TCCTTATTTT TAAATAAACT CTCTAACTCC TCATAGCTAT TAACTGTCAA
ATCTCCACCA AATTTTTCTG GCCTTTTTAT GGTTAAAGGA TATTCAAGGA
AGTATTTAGC TATCTCCATT ATTGGATTTC CTTCAACAAC TCCAGCTGGG
CAGTATGCTT TCTTTATCTT AGCCCTAATC TCTTCTGGAG AGTCATCAAC
AGCTATAAAA TTCCCTTTTG AAGAACTCAT CTTTCCTTCT CCATCCAAAC
CCGTTAAGAC AGGGTTGTGA ATACAAACAA CCTTTTTTGG TAAAAGCTCC
CTTGCTAACA TGTGTATTTT TCTCTGCTCC ATCCCTCCAA CTGCAACATC
AACGCCTAAA TAATGAATAT CATTAACCTG CATTATTGGA TAGATAACTT
CAGCAACCTT TGGATTTTCA TCCTCTCTTG CTATAAGTTC CATACTCCTT
CTTGCTCTTT TTAAGGTAGT TTTTAAAGCC AATCTATAGA CATTCAGTGT
ATAATCCTTA TCAAGCTGGA ATTCagcgtt acaagtatta cacaaagttt
tttatgttga gaatattttt ttgatggggc gccacttatt tttgatcgtt
cgctcaaagA AGCGGCGCCA GGGNTGTTTT TCTTTTCACC AGTNAGACGG
GCAACAGAAC GCCATGAgcg gcctcatttc ttattctgag ttacaacagt
ccgcaccgct gtccggtagc tccttccggt gggcgcgggg catgactatc
gtcgccgcac ttatgactgt cttctttatc atgcaactcg taggacaggt
gccggcagcg cccaacagtc ccccggccac ggggcctgcc accataccca
cgccgaaaca agcgccctgc accattatgt tccggatctg catcgcagga
tgctgctggc taccctgtgg aacacctaca tctgtattaa cgaagcgcta
accgttttta tcaggctctg ggaggcagaa taaatgatca tatcgtcaat
tattacctcc acggggagag cctgagcaaa ctggcctcag gcatttgaga
agcacacggt cacactgctt ccggtagtca ataaaccggt aaaccagcaa
tagacataag cggctatta acgaccctgc cctgaaccga cgaccgggtc
gaatttgctt tcgaatttct gccattcatc cgcttattat cacttattca
```

		ggcgtagcac caggcgttta agggcaccaa taactgcctt aaaaaaatta cgccccgccc tgccactcat cgcagtactg ttgtaattca ttaagcattc tgccgacatg gaagccatca cagacggcat gatgaacctg aatcgccagc ggcatcagca ccttgtcgcc ttgcgtataa tatttgccca tggtgaaaac gggggcgaag aagttgtcca tattggccac gtttaaatca aaactggtga aactcaccca gggattggct gagacgaaaa acatattctc aataaaccct ttagggaaat aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaaacgtttc agtttgctca tggaaaacgg tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtctttcatt gccatacg
SEQ ID: 82	pKQ	ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT GGTAGTGTGG GGTCTCCCCA TGCGAGAGTA GGGAACTGCC AGGCATCAAA TAAAACGAAA GGCTCAGTCG AAAGACTGGG CCTTTCGTTT TATCTGTTGT TTGTCGGTGA ACGATATCTG CTTTTCTTCG CGAATtaatt ccgcttcgca ACATGTgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg TCATGAgttg tgtctcaaaa tctctgatgt tacattgcac aagataaaaa tatatcatca tgaacaataa aactgtctgc ttacataaac agtaatacaa ggggtgttat gagccatatt caacgggaaa cgtcttgctc gaggccgcga ttaaattcca acatggatgc tgatttatat gggtataaat gggctcgcga taatgtcggg caatcaggtg cgacaatcta tcgattgtat gggaagcccg atgcgccaga gttgtttctg aaacatggca aaggtagcgt tgccaatgat gttacagatg agatggtcag actaaactgg ctgacggaat ttatgcctct tccgaccatc aagcatttta tccgtactcc tgatgatgca tggttactca ccactgcgat ccccgggaaa acagcattcc aggtattaga agaatatcct gattcaggtg aaaatattgt tgatgcgctg gcagtgttcc tgcgccggtt gcattcgatt cctgtttgta attgtccttt taacagcgat cgcgtatttc gtctcgctca ggcgcaatca cgaatgaata acggtttggt tgatgcgagt gattttgatg acgagcgtaa tggctggcct gttgaacaag tctggaaaga atgcataag cttttgccat tctcaccgga ttcagtcgtc actcatggtg atttctcact gataaccctt attttgacg aggggaaatt aataggttgt attgatgttg acgagtcgg aatcgcagac cgataccagg atcttgccat cctatggaac tgcctcggtg agttttctcc ttcattacag aaacggcttt ttcaaaaata tggtattgat aatcctgata tgaataaatt gcagtttcat ttgatgctcg atgagttttt ctaatcagaa ttggttaatt ggttgtaaca ctggcagagc attacgctga cttgacggga cggcggcttt gttgaataaa tcgaactttt gctgagttga aggatcCTCG GGagttgtca gcctgtcccg cttataagat catacgccgt tatacGTTGT TTACGCTTTG AGGAATTAAC C